

# Implementing the START:AV in a Dutch Residential Youth Facility

Citation for published version (APA):

De Beuf, T. L. F., De Vogel, V., & De Ruiter, C. (2019). Implementing the START:AV in a Dutch Residential Youth Facility: Outcomes of Success. *Translational Issues in Psychological Science*, 5(2), 193-205. <https://doi.org/10.1037/tps0000193>

## Document status and date:

Published: 01/06/2019

## DOI:

[10.1037/tps0000193](https://doi.org/10.1037/tps0000193)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

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

[Link to publication](#)

## General rights

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

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

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

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

## Take down policy

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

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

providing details and we will investigate your claim.

## Implementing the START:AV in a Dutch Residential Youth Facility: Outcomes of Success

Tamara L. F. De Beuf

OG Helderling Institution, Zetten, the Netherlands,  
and Maastricht University

Vivienne de Vogel

University of Applied Sciences Utrecht and The  
Forensic Care Specialists, Utrecht, the Netherlands

Corine de Ruiter

Maastricht University

Currently, implementation research in the field of forensic risk assessment is limited and consensus on “implementation success” is lacking. This study applies outcomes of success from implementation science to the implementation evaluation of the Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV) in a residential youth care facility in the Netherlands. Staff perceptions on the implementation and the instrument were assessed using 5 implementation outcomes in a longitudinal multimethod design: acceptability, adoption, appropriateness, feasibility, and penetration. As anticipated, the majority of staff perceived START:AV core constructs as useful for treatment (appropriateness). However, satisfaction with the instrument decreased over time (acceptability). This was likely due to an increased workload (feasibility). Despite this dissatisfaction, the completion rate was acceptable (adoption). Lastly, staff reported a lack of integration of the START:AV findings in clinical case conferences (penetration). The implementation outcomes aid in identifying areas for improvement, which in turn can lead to an increased and more consistent uptake of structured risk assessment into routine practice.

### *What is the significance of this article for the general public?*

Implementation science defines several outcomes of implementation success, such as acceptability, adoption, and appropriateness. We used five outcomes to evaluate staff perceptions on the implementation of the START:AV risk assessment tool in a residential youth care facility. The findings on the outcomes provided extensive information on the implementation progress and generated valuable clues for increasing the quality of the risk assessment implementation.

**Keywords:** START:AV, implementation, implementation outcomes, risk assessment, residential youth care

**Supplemental materials:** <http://dx.doi.org/10.1037/tps0000193.supp>

Tamara L. F. De Beuf, OG Helderling Institution, Zetten, the Netherlands, and Department of Clinical Psychological Science, Maastricht University; Vivienne de Vogel, Research Group Working With Mandated Clients, University of Applied Sciences Utrecht, and Department of Research, The Forensic Care Specialists, Utrecht, the Netherlands; Corine de Ruiter, Department of Clinical Psychological Science, Forensic Psychology Section, Maastricht University.

The OG Helderling Institution publishes the START:AV user guide and provides training in the instrument. All proceeds are used to fund research.

Correspondence concerning this article should be addressed to Tamara L. F. De Beuf, Ottho Gerhard Helderlingstichting, P.O. Box 1, Zetten 6670 AA, the Netherlands. E-mail: [t.debeuf@oghelderling.nl](mailto:t.debeuf@oghelderling.nl)

The field of violence risk assessment research is rapidly growing. Various systematic approaches to risk assessment have emerged: the actuarial and structured professional judgment (SPJ) approach (Brown & Singh, 2014). Both rely on a set of risk and/or protective factors that have empirical support as to their effect on the likelihood that an adverse event will occur (Brown & Singh, 2014). The approaches differ in how they reach a conclusion on a person's risk of recidivism. Briefly, the actuarial instruments rely on an algorithm, to derive a probabilistic estimate of the likelihood that the adverse outcome (e.g., violence, sexual recidivism) will occur. SPJ instruments leave more to the professional's discretion, while simultaneously providing definitions and item indicators as formal guidance to the risk assessment. The assessor weights, combines and integrates the risk factors to determine whether the risk is low, moderate or high (Nicholls, Petersen, & Pritchard, 2016).

Both actuarial and SPJ approaches generated multiple risk assessment instruments (Singh et al., 2014) and it is estimated that over a million structured risk assessments are conducted annually around the world (Viljoen, Cochrane, & Jonnson, 2018). The application of risk assessment tools originates from a widespread belief that they can assist in managing and reducing adverse outcomes such as reoffending (Viljoen et al., 2018). Still, there is a lack of empirical evidence that risk assessment instruments are effective in reducing risk (Desmarais, 2017). A recent systematic review by Viljoen and colleagues (2018) found insufficient evidence to conclude that risk assessment tools directly reduce violence and offending or that they enhance professionals' risk management practices. Notwithstanding these findings, the authors state that structured risk assessment remains a best available practice to date. They formulate recommendations for research and practice to enhance the potential value of risk assessment instruments. Increasing implementation quality by following up on the adherence to the instrument's instructions, is one of their suggestions to improve the utility of structured risk assessment. For example, Vincent, Guy, Gershenson, and McCabe (2012) found that probation officers did not incorporate level of risk in their decision-making until relevant policies were in place, new case management

forms were implemented and they were trained in using risk assessment in their decisions.

### Implementation Science

Implementation issues are not exclusive to the field of risk assessment. Incorporating evidence-based methods into routine practice is a common challenge in (mental) health care (Proctor et al., 2009). It is estimated that it takes almost two decades for new health care interventions to become integrated into real-world contexts (Morris, Wooding, & Grant, 2011). This gap between research and practice prompted the rise of implementation science. Implementation science is the study of methods to promote the uptake of evidence-based methods into routine practice, with the aim of improving health services (Eccles & Mittman, 2006). Implementation research addresses three features of the implementation process: (a) factors affecting the implementation (implementation determinants), (b) strategies to deliver the implementation (implementation strategies), and (c) the results of the implementation (implementation outcomes; Peters, Tran, & Adam, 2013). This article will focus on the latter.

### Implementation Outcomes

Implementation outcomes are the results of intentional actions to implement a new evidence-based practice (Proctor et al., 2009). They serve as a precondition for obtaining change in effectiveness outcomes (e.g., reduction in violent incidents; Proctor et al., 2009). The World Health Organization (Peters et al., 2013) recommends the explicit use of implementation outcomes to assess the adequacy of an implementation. Equally important, the use of a common set of implementation outcomes enables comparative research (Proctor et al., 2011). Proctor and colleagues (2011) developed a taxonomy of eight implementation outcomes: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability (see Table 1). This taxonomy, one of the most cited frameworks in implementation science, is referred to as the Implementation Outcomes Framework (IOF; Lewis et al., 2015). The authors acknowledge dynamic and complex interrelations between the outcomes.

Table 1  
*Implementation Outcomes Framework*

Implementation outcome	Working definition	Related terms
Acceptability	The perception among stakeholders that a given risk assessment instrument is agreeable based on content, complexity, convenience, and credibility	Satisfaction
Adoption	The intention, initial decision, or action to try a new risk assessment instrument	Uptake, utilization, intention to try
Appropriateness	The perceived fit or relevance of the risk assessment instrument for a given setting or population	Perceived fit, relevance, compatibility, suitability, usefulness, practicability
Feasibility	The extent to which a risk assessment instrument can be accurately carried out within a given setting or organization	Actual fit, utility, suitability for everyday use, practicability
Fidelity	The degree to which a risk assessment instrument is carried out as it was designed in the original protocol or manual	Adherence, integrity, delivery as intended, quality of program delivery
Implementation costs	The cost of the effort to implement a new risk assessment instrument	Marginal cost, cost–benefit, cost-effectiveness
Penetration	The degree to which a risk assessment instrument is integrated into a service setting	Coverage (however, this refers to the degree to which a population actually receives an intervention)
Sustainability	The extent to which a newly implemented risk assessment instrument is maintained	Maintenance, durability, continuation, incorporation, integration, institutionalization

*Note.* Adapted and applied to risk assessment from “Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda” by E. Proctor et al. (2011), *Administration and Policy in Mental Health and Mental Health Services Research*, 38, p. 68. Copyright 2010 by the Authors.

### Research on Risk Assessment Implementation

Studies on the implementation of violence risk assessment tools are scarce. A systematic review identified 11 studies analyzing the implementation of structured risk assessment instruments in psychiatric and correctional settings (Levin, Nilsen, Bendtsen, & Bulow, 2015). The authors applied the Consolidated Framework for Implementation Research (Damschroder et al., 2009) to discuss reported implementation barriers and facilitators. To our knowledge, this was the first effort to use a conceptual framework from the field of implementation science in forensic risk assessment research. The authors identified a variety of definitions of “implementation success” across the reviewed studies. Some studies inferred success from the ongoing maintenance of the risk assessment practice over time (Crocker et al., 2011; Kroppan et al., 2011) or from the hospital-wide diffusion of the risk assessment practice (Vojt, Slessor, Marshall, & Thomson, 2011), whereas others inferred success from

user convenience and positive responses from staff (Clarke, Brown, & Griffith, 2010). The majority of reviewed studies did not describe their implementation outcome. Without defining implementation outcomes, it is difficult to accurately infer the effectiveness of the implementation and to make suggestions for future implementation processes in forensic risk assessment. Therefore the present study includes outcomes from implementation science to evaluate a risk assessment implementation endeavor.

### The Present Study

The goal of the present study is to evaluate staff perceptions of the implementation of the Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV; Viljoen, Nicholls, Cruise, Desmarais, & Webster, 2014) in a residential youth care setting in the Netherlands. Five outcomes from the IOF (Proctor et al., 2011) were used: acceptability, adoption, appropriateness, feasibility, and penetration. Implementation costs and sustainability were

not included, as they could not be assessed within the study's timeframe. Fidelity will be reported in a separate article. For each implementation outcome, hypotheses were formulated. We first describe the implementation process in Figure 1 and the implemented instrument in the Method section.

## Method

### Setting

The study was conducted at one of 14 centers in the Netherlands that provide residential youth care mandated by court. Mandated residential

youth care is the most restrictive form of residential treatment for adolescents who suffer from severe behavioral and mental health problems. A high level of supervision is required to guarantee their safety (e.g., suicidal behavior or sexual exploitation) and/or the safety of their environment (e.g., physical violence toward others). Mandated residential youth care is distinct from juvenile detention. A major difference is that adolescents are admitted under civil law with a child protection order, whereas in juvenile detention, youths are admitted under juvenile criminal law. At the time of the study, the facility had a capacity of 98 beds: three

The START:AV implementation was based on guidelines presented in *Risk Assessment in Juvenile Justice: A Guidebook for Implementation* by Vincent, Guy, and Grisso (2012). We summarize the implementation process according to the guidebook's eight steps.

1. *Get Ready.* The institution connected with a risk assessment expert (third author), hired a project coordinator (PC; first author), and established an implementation committee consisting of the clinical director, local researcher, two treatment coordinators, and the PC. Staff learned about the project during a symposium and was kept informed via timely updates in newsletters.
2. *Establish staff buy-in.* Buy-in was prompted by consulting staff during the decision-making process and educating them on the instrument and its implications for practice.
3. *Select a tool.* The START:AV was considered the most appropriate validated risk assessment instrument for this setting and population. In contrast to other risk assessment instruments, it is designed to assess the risk of multiple adverse outcomes, which are all relevant to the youth in residential care (Vermaes, Konijn, Jambroes, & Nijhof, 2014). Another unique feature is the tool's balanced approach to risk and protective factors, which corresponds with the setting's philosophy to observe, validate and strengthen the assets of a youth. Moreover, because of its dynamic nature, the instrument can detect reliable change over short time periods, which is in line with the setting's treatment cycle of 4 months (Sellers, Desmarais, & Hanger, 2017). The START:AV forms and user guide were officially translated into Dutch (De Beuf, de Ruiter, & de Vogel, 2016), with permission from and in consultation with the original authors.
4. *Prepare policies.* The PC and the implementation committee prepared policies for the use and the integration of the tool within the treatment cycle and plans. For example, a risk formulation section ("Dynamic Risk Profile") was added to the treatment plan, treatment goals were linked to the items, and a START:AV timeline was developed summarizing the crucial stages of an assessment.
5. *Provide training.* Treatment coordinators participated in a 2-day training on the use of the START:AV for risk assessment and intervention planning and completed additional practice cases. Newly hired treatment coordinators received one-on-one training shortly after recruitment. All training was provided by the first author.
6. *Implement pilot test.* A pilot implementation was carried out on two living groups from June until November 2015 to evaluate the adjusted documents and policies and the START:AV assessment process. The PC discussed the findings with the implementation committee and carried out subsequent recommendations.
7. *Begin full implementation.* The official implementation started on February 1, 2016, simultaneously on all units, for all newly admitted adolescents.
8. *Follow up.* The PC wrote internal reports on evaluation moments (T1 and T2 in this study) and discussed the findings with the implementation committee and treatment coordinators. The PC carried out follow-up actions to address the encountered challenges. For example, booster sessions for treatment coordinators and information sessions for other staff were organized following each survey. The PC was readily available for consultation.

Figure 1. Implementation of the START:AV in a residential youth care setting.



high-secure (33) and six medium-secure units (65). In 2016, 137 adolescents were admitted for an average duration of 8.6 months. Their average age was 15.5 years old and 52% were girls.

The setting's primary motivation to adopt structured risk assessment was to monitor and manage the youth's risks. In mandated residential care, the task is to reduce risk so that adolescents can be transferred to a less restrictive environment for further care. Structured risk assessment was deemed necessary for this purpose. Secondary goals of risk assessment were to inform treatment decisions and reduce treatment duration and incidents.

### START:AV

The START:AV is a risk assessment instrument for boys and girls aged 12 to 18 years. It is designed to assess short-term risk of multiple outcomes, including violence to others, nonviolent offending, substance abuse, unauthorized absence, suicide, self-harm, victimization, and self-neglect. The instrument comprises 24 dynamic items that are associated with these adverse outcomes. Each item is rated twice on a 3-point scale (low, moderate, high): once as a strength and once as a vulnerability, based on the past 3 months. As an SPJ tool, the START:AV does not result in a total score that automatically indicates the risk level. The assessor formulates a final risk judgment (Low, Moderate, High risk) for all adverse outcomes separately. The START:AV's psychometric properties indicate fair to excellent interrater reliability (intraclass correlation coefficient,  $s = .52$  to  $.92$ ) and significant predictive validity for the adverse outcomes (areas under the curve =  $.63$  to  $.83$ ; Bhanwer, Shaffer, & Viljoen, 2016).

**START:AV practice.** The START:AV user guide (Viljoen et al., 2014) mentions an administration time of approximately 30 min. This estimation merely concerns the time required for rating the items and adverse outcomes. The present setting adopted the START:AV comprehensive rating form and considered it as a "master file" because all available information about the youth (e.g., psychological testing, observations, file information) was gathered in this form. Treatment coordinators (licensed psychologists) completed the START:AV assessments. Initially, they were the only staff members reporting in the START:AV, however, 1 year into the implemen-

tation, the management decided that all staff involved with the adolescent (e.g., therapist, group care workers) had to report directly into the START:AV form. In particular group care workers, as frontline staff on the adolescent's unit, contributed extensively to the documentation on the START:AV items. The treatment coordinators remained responsible for rating the items and providing final risk judgments. A START:AV assessment was conducted at least every 4 months, prior to each new or revised treatment plan.

### Procedure

The study was part of a larger qualitative and quantitative implementation evaluation. For the present article, we used data gathered with two quantitative measures, a file audit and a user survey. The qualitative findings will be reported elsewhere. Because the implementation of a structured risk assessment tool is a process that unfolds over time, a longitudinal design was considered appropriate. The study covered 17 months, from March 2016 until July 2017, with two follow-ups. The first (T1) was at 7 months and the second (T2) at 17 months.

### Participants

The study focused exclusively on the perceptions and output of treatment coordinators as they were ultimately responsible for the START:AV assessments. Due to staff changes during the 17-month period, there were 16 unique treatment coordinators (14 and two trainees) who completed START:AVs. At T1, the team consisted of 10 treatment coordinators of which eight completed the survey (response rate = 80%). At T2, there were nine practicing treatment coordinators of which seven responded (response rate: 78%), however one treatment coordinator did not complete the survey. The responding staff members were all female and the majority (60%) was between 35 and 44 years old. Forty percent of the respondents had more than 10 years of service within the institution, 33% had between 5 and 10 years of service, and 27% had 1 year or less. They all held an academic degree.

### Hypotheses

The hypotheses of the present study were primarily informed by a prior implementation of the START:AV in a medium secure adoles-

cent service in the United Kingdom (Sher & Gralton, 2014). In this study, the START:AV was completed every 3 months through consensus of at least three trained professionals involved with the assessed adolescent. Six months into the implementation, the authors surveyed the multidisciplinary teams on their views about the implementation. Staff reported overall satisfaction with the instrument. They indicated that it was straightforward to assess strengths and vulnerabilities; however, about half of the surveyed staff experienced some difficulties distinguishing between the item ratings (i.e., low, moderate or high). Staff highlighted that the quality of the START:AV largely depended on familiarity with the adolescent and the majority agreed that the required information was readily available. Overall, the START:AV was viewed as applicable to their service users, providing a comprehensive risk assessment. Staff particularly supported the dynamic nature of the assessment as well as the balanced focus on strengths and vulnerabilities. They felt that rating the START:AV as a team improved communication, teamwork and contributed to a deeper understanding of the service users. Cited challenges centered on time effort and increased workload. In light of these findings, we formulated the following hypotheses:

*Hypothesis 1: Acceptability.* Similar to Sher and Gralton (2014), we expected staff to be satisfied with the instrument, albeit with some dissatisfaction due to an increased workload. We expected increased acceptance over time, in line with Proctor et al.'s (2011) assumption that acceptability changes with experience.

*Hypothesis 2: Adoption.* We expected an overall uptake of the START:AV of at least 80%, because considerable attention was paid to embedding the instrument in routine practice. We also anticipated lower completion rates for some treatment coordinators who would be less inclined to complete the START:AV. This assumption is based on Rogers's (1962) theory of diffusion of innovations, which identifies fast and slow adopters of an innovation.

*Hypothesis 3: Appropriateness.* The START:AV was expected to be perceived as compatible with the setting's philosophy due to its focus on dynamic risk factors, its in-

clusion of strengths and its relevance for case management (see Figure 1).

*Hypothesis 4: Feasibility.* In line with Sher and Gralton (2014), we expected that staff would consider it feasible to complete a START:AV assessment with the available information, albeit with criticism about an increased workload. This challenge is mentioned in various risk assessment implementation studies (Levin et al., 2015).

*Hypothesis 5: Penetration.* We predicted sufficient integration of the START:AV assessments in the setting's treatment plans and case conferences. We expected that, over time, staff would report increasing penetration.

## Measures

**START:AV records.** Treatment coordinators completed START:AV forms in Microsoft Word and saved them on a shared drive, which was available to the first author. The first author monitored completed and missing START:AV forms using a START:AV tracking file in Excel. Each month, she sent out a reminder to all treatment coordinators about missing forms and planned assessments. A START:AV was considered missing when a scheduled assessment was not conducted.

**START:AV user survey.** The START:AV user survey was translated and adapted from the START:AV user satisfaction survey by Sher and Gralton (2014). The adjusted version is a 2.5-page survey, administered via Qualtrics software (mean completion time = 6 min). The survey begins with questions on staff demographics and START:AV experience, and continues with staff need for additional training. The body of the survey contains 34 statements about (the use of) the instrument, for example, "I had sufficient time to rate the START:AV," or more general, "The focus on dynamic factors is useful for treatment." All statements are rated on a 5-point Likert scale from *strongly disagree* to *strongly agree*. The survey was filled out voluntarily and anonymously on both time points.

## Ethical Considerations

Permission to conduct the research was obtained from the institution's General Director.

The Ethics Review Committee Psychology and Neuroscience (ERCPN) of Maastricht University approved the research protocol and consent forms (ERCPN Number 174\_06\_12\_2016). All data were analyzed anonymously and stored according to the university's data management code of conduct and the institution's data protection guidelines, based on Dutch and European legislation.

### Data Analysis

Using SPSS Version 25, a frequency analysis was conducted on the START:AV records for two time periods: March to September 2016 and October 2016 to July 2017. All data from the START:AV survey were transferred from Qualtrics into SPSS. For the purpose of analysis, the response categories 'strongly agree' and 'agree' were collapsed as well as 'strongly disagree' and 'disagree'. The distribution of responses on the survey statements was computed with a frequency analysis and the Somers' *d* statistic was calculated to assess whether the distribution changed significantly over time. Somers' *d*

measures the strength and direction of an association between two ordinal variables (i.e., time and statement). However, because the surveys were completed anonymously, we could not assess whether changes over time resulted from differences in sample composition. Table 2 presents the measures, analyses and main findings.

### Results

Tables with the response distribution (*agree, more or less, disagree*) on the survey statements are available from the first author upon request as well as in the [online supplemental materials](#).

#### Adoption: How Frequently Are START:AV Assessments Completed by the Users?

During the first 7 months, 12 assessors completed 59 START:AVs for 53 unique youths. On average, 8.4 START:AVs were completed monthly. During the second time period, 13 assessors completed 140 START:AVs for 101

Table 2  
Overview of the Assessed Implementation Outcomes With Used Measures, Applied Analysis, and Main Findings

Implementation outcome	Measure	Analysis	Principal findings	Significant changes over time
Acceptability	START:AV User Survey	Frequency analysis; Somers' <i>d</i> statistic	Dissatisfied users at Time (T)2; overall limited credibility but considerable confidence among users at T1 and T2	Negative correlation for Time × Easily Distinguish Between Low/Moderate/High; Negative Correlation for Time × Overall Satisfaction With START:AV
Adoption	START:AV Records	Frequency analysis	Average completion rate of 74% at T1 and 78% at T2; individual completion rates ranging from 29% to 100%	n/a
Appropriateness	START:AV User Survey	Frequency analysis; Somers' <i>d</i> statistic	Agreed usefulness of START:AV key components, not general usefulness	Negative correlation for Time × START:AV as a Useful Tool; Negative Correlation for Time × Useful to Rate Vulnerabilities
Feasibility	START:AV User Survey	Frequency analysis; Somers' <i>d</i> statistic	Lack of time to complete assessments	Negative Correlation for Time × Time Needed for Rating the Items
Penetration	START:AV User Survey	Frequency analysis; Somers' <i>d</i> statistic	Insufficient integration at T1; no increase in effective communication or in structure of meetings at T1 and T2	Positive Correlation for Time × START:AV Is Sufficiently Integrated

Note. START:AV = Short-Term Assessment of Risk and Treatability: Adolescent Version; n/a = not applicable.



unique youths. This resulted in an average of 14 START:AVs per month. The minimum completion rate required by the service's management was set at 80%. During the first period, 21 START:AVs were missing, resulting in a completion rate of 74%. During the second period, the completion rate was 78%, with 38 missing. As can be seen in Figure 2, the average completion rate varied among assessors, ranging from 28.6% (Assessor 11) to 100% (Assessors 2 and 13).

**Acceptability: How Satisfied Are Users With the START:AV?**

On the statement about overall satisfaction with the instrument, assessors increasingly indicated they were dissatisfied with the START:AV, with none of the assessors dissatisfied at T1 and 66.6% dissatisfied at T2. The strong, negative correlation between time and satisfaction was statistically significant (Somers'  $d = -.67, p = .005$ ). Other components of acceptability relate to content, credibility, and complexity of the instrument. With regard to content, at both times, the majority of assessors agreed that the START:AV provides a complete picture of the adolescent, with 100% agreeing at T1 and 66.6% at T2. With regard to credibility, only a limited percentage of the assessors believed the START:AV would help reduce incidents within the setting (T1 = 25.0%; T2 = 16.7%). The

complexity of the risk assessment instrument was not an issue for staff. Most assessors agreed, at least more or less, that they could accurately assess strengths and vulnerabilities independently of each other (T1 = 87.5%; T2 = 83.3%). However, over time, more assessors seemed to find it difficult to distinguish between "low," "moderate," and "high" ratings (Somers'  $d = -.65, p = .007$ ); the percentage of assessors who disagreed that they could easily distinguish between the ratings increased from 12.5% at T1 to 60.0% at T2. Still, at T2, 60% agreed they were confident in their item ratings and none of the assessors communicated a training need concerning the items. Although one or two assessors did not feel entirely confident about their final risk judgment on suicide, self-harm and victimization, all assessors agreed to be confident about their risk estimates on the other adverse outcomes. Seventeen months into the implementation, all assessors indicated that they felt confident in their ability to complete the START:AV and in the accuracy of their judgments.

**Appropriateness: How Useful Is the START:AV According to Users?**

The majority of the assessors (T1 = 62.5%; T2 = 66.7%) agreed more or less that the START:AV is useful for treatment. However, over time, there was a significant shift with

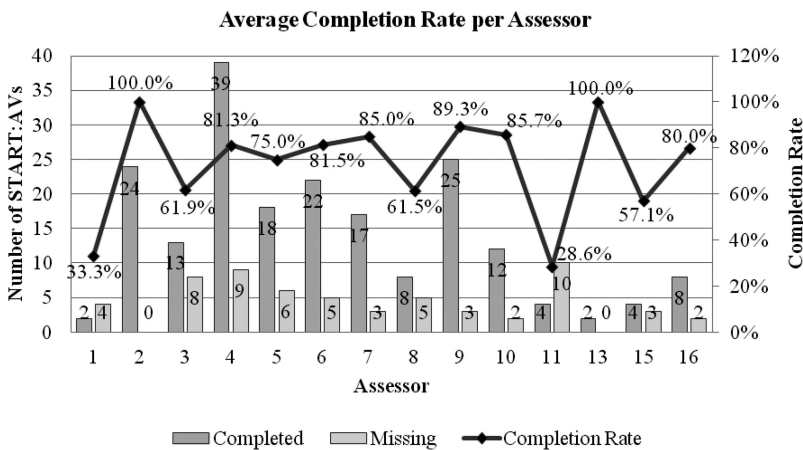


Figure 2. The number of completed and missing Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV) assessments and completion rates per assessor. Assessors 12 and 14 were excluded from the graph as they were trainees without a caseload of their own. They completed START:AVs from the caseload of other assessors.

more assessors disagreeing with usefulness (Somers'  $d = -.58, p = .001$ ); no assessors disagreed at T1 but 33.0% did at T2. Still, a large majority of assessors agreed that the focus on dynamic factors was useful for treatment (T1 = 75.0%; T2 = 100%) and agreed that combining strengths and vulnerabilities was useful (100% at T1 and T2). In addition, a large majority of assessors found identifying key strengths and critical vulnerabilities useful (T1 = 85.7%, T2 = 100%; and T1 = 85.7%, T2 = 80.0%, respectively). Likewise, more than three-quarters of the assessors agreed at both times that the risk estimates and the dynamic risk profile were valuable. On the other hand, not all assessors were convinced of the benefits of rating vulnerabilities; there was a downward trend in the percentage of assessors who fully agreed with the usefulness of rating vulnerabilities (Somers'  $d = -.50, p = .051$ ). Lastly, the option to document a personal risk signal was found useful by slightly more than half of the assessors (T1 = 57.1%, T2 = 60.0%).

### **Feasibility: How Practical Is the START:AV Assessment According to Users?**

A lack of time to complete a START:AV assessment was reported by most treatment coordinators (T1 = 75%; T2 = 100%). To gain insight into the workload related to the START:AV, we asked treatment coordinators to estimate how much time they spent completing the form. Seven months into the implementation, assessors reported that a START:AV assessment took 3 hr on average to complete, including 1 hr to rate the items. Ten months later, it took on average between 2 and 3 hr, including 15–30 min to rate the items. The time needed to rate the items decreased significantly over time (Somers'  $d = -.63, p = .002$ ). Furthermore, most assessors agreed, at least more or less, that the required information was readily available (T1 = 87.5%, T2 = 100%) and agreed they knew the adolescent well enough to complete the START:AV (T1 = 75%, T2 = 83.3%).

### **Penetration: How Integrated Is the START:AV Within the Service According to Users?**

At both T1 and T2, nearly 100% of assessors disagreed that the START:AV contributed to

more effective communication and that it increased structure during case conferences. On the other hand, over time, the integration into the treatment process seemed to improve as the percentage of assessors who disagreed that the START:AV was sufficiently integrated decreased from 75% at T1 to 17% at T2 (Somers'  $d = .50, p = .048$ ).

## **Discussion**

To our knowledge, this is the first empirical study that evaluates the implementation of a structured risk assessment tool using outcomes from implementation science. Five outcomes in particular served as indicators for implementation success: adoption, acceptability, appropriateness, feasibility, and penetration.

Contrary to our first hypothesis, the assessors displayed overall dissatisfaction with the instrument, with a sharp increase in dissatisfaction over time. This might be explained by the assumption of Proctor et al. (2011) that feasibility can affect acceptability. That is, an increase in the treatment coordinators' workload likely contributed to the reported discontent. The comprehensive assessment approach led to a higher workload due to considerable copy-and-pasting from files and reports into the START:AV rating form. In addition, assessors spent much time documenting arguments for all 24 START:AV items. To address these challenges, work processes were adjusted in between surveys and other professionals, such as teachers and social workers, started reporting directly into the START:AV form, resulting in less copy-and-pasting for treatment coordinators. Still, the desired reduction in workload was not evident at T2. The absence of the anticipated workload reduction might in fact have increased the dissatisfaction with the instrument over time. Although the decrease in satisfaction was unexpected, it supports Proctor et al.'s (2011) assumption that acceptability is a dynamic construct. Nevertheless, the board of directors insisted on continuing the implementation; they engaged in a dialogue with staff and offered solutions such as digitizing the rating form.

Regarding the adoption hypothesis, we found a slightly lower uptake than the targeted 80%. Proctor and colleagues (2011) suggested that adoption is likely to be affected by acceptability. Thus, assessors who are less satisfied are

more likely to demonstrate lower adoption rates. Considering the relatively low satisfaction among users, particularly at T2, a lower adoption was expected. Although the uptake was below the management's objective of 80%, it was still acceptable. The completion rate varied considerably among assessors and seemed to fit the adopter categories originating from the theory of diffusion of innovations (Rogers, 1962). Early adopters (Assessors 2 and 13), early majority (Assessors 4, 6, 7, 9, and 10), late majority (Assessors 3, 5, 8, 15, and 16), and laggards (Assessors 1 and 11) could be identified. Due to the anonymous nature of the survey, we could not examine whether the "laggards" were the assessors who showed less acceptability. In response to the suboptimal adoption, the setting's management decided to monitor START:AV completion rates every 2 months. Assessors with a low completion rate were approached and their need for support was explored. The monitoring effort resulted in an improved completion rate (95% in 2018), with the majority of assessors completing 100% of their START:AVs. In line with Roger's theory, the late majority and laggards likely needed more time and, perhaps, pressure to adopt the instrument. Also, some staff members with a lower completion rate left the facility.

The appropriateness hypothesis was supported in part: assessors perceived the START:AV core components (i.e., dynamic factors, strengths, key and critical items, risk estimates, and the dynamic risk profile) as useful for treatment. Similar findings were reported by Sher and Gralton (2014). Still, over time, more assessors disagreed with the overall usefulness of the instrument. Although Proctor and colleagues (2011) do not suggest a potential effect of acceptability on appropriateness, we contemplate that dissatisfaction may have colored the perception of usefulness. This issue was addressed by reaffirming the rationale for conducting structured risk assessments via a management statement.

As expected with regard to feasibility, staff felt that there was insufficient time for the assessments. The average completion time was about six times higher than reported in the U.K. study (i.e., 3 hr vs. 27 min; Sher & Gralton, 2014). However, two differences in approach should be mentioned. First, in the U.K. implementation, professionals only wrote down brief prompts while rating (M. Sher, personal com-

munication, November 11, 2015), whereas in the present setting, treatment coordinators were instructed to report all available and relevant information in the START:AV form to motivate their ratings. Second, in the U.K. study, assessors were expected to be familiar with the youth's file prior to the assessment, while in the present study, reading the available documentation was also captured in the completion time. In accordance with our expectation and previous studies, assessors became more efficient over time; the average time to rate the items decreased from 1 hr to 15–30 min. As mentioned earlier, treatment coordinators' objections about time investment were taken seriously and pros and cons of alternative completion scenarios were discussed, ranging from omitting item ratings and merely rating the adverse outcomes, to adhering to the master-file approach. The team of treatment coordinators agreed the latter scenario was most relevant and useful for treatment. However, they abandoned detailed motivations to some extent.

Contrary to our hypothesis on penetration, only a small proportion felt that there was sufficient integration of the START:AV in multidisciplinary case conferences. Still, more staff observed integration over time. However, the positive effect of the START:AV on communication, teamwork and planning as reported by Sher and Gralton (2014), was not replicated in the present setting. These benefits possibly resulted from the consensus approach the authors applied. That is, each START:AV was rated through consensus by three professionals directly involved with the adolescent, which was likely more conducive to teamwork than the rating by a single professional in our study. Although we contemplated the consensus model, it was deemed not feasible to organize multidisciplinary START:AV meetings within the service. Subsequent efforts to increase penetration included enhanced incorporation of risk information in treatment plans, inclusion of case alerts (based on high-risk estimates from the START:AV) in the adolescent's electronic record, and the option to label progress notes with specific START:AV items. Furthermore, treatment coordinators received suggestions on how to further integrate the START:AV into case conferences.

## Limitations

A major limitation of the study is the small sample of assessors. However, the response rate on the survey was almost 80%, representing the majority of assessors. The anonymous completion of the survey is a second limitation because this impeded the interpretation of demonstrated changes over time. A third limitation is the utilization of an existing survey with unknown psychometric properties. On the one hand, this allowed direct comparison with a previous START:AV implementation study (Sher & Gralton, 2014). On the other hand, the survey did not fully cover the IOF outcomes. Recently, Weiner et al. (2017) developed promising open-access measures for acceptability, appropriateness, and feasibility, which might be relevant for future studies. In addition, psychometrically sound measures would allow for the modeling of interrelations between implementation outcomes (Proctor et al., 2011). A final limitation is that we did not survey all relevant stakeholders, for example, perceptions of management and youths were not examined. Their attitudes can facilitate or hinder the implementation process and are important to include in future evaluations (Levin et al., 2015; Müller-Isberner, Born, Eucker, & Eusterschulte, 2017). During the present implementation, the managing director, who was a strong advocate for structured risk assessment, left the service, resulting in more fragile support for risk assessment at the managerial level. This might have impacted the implementation process (see also Kroppan et al., 2011). In future implementation studies, youth and youth's support system could be surveyed about their experiences with risk assessment implementation (Proctor et al., 2011).

## Practical Implications

Findings at each time point were presented to the implementation committee, management and treatment coordinators (see Figure 1) and generated improvement strategies that were discussed and carried out. The implications for the present setting, as reported in the discussion, demonstrate the importance of staff involvement, monitoring, software assistance, top-down support and perseverance. Crucial to the implementation persistence was that management learned about the change curve of Kübler-

Ross (1969), initially developed to explain a person's grieving process and later utilized as a change management model (Belyh, 2015). This change curve depicts the emotional stages that individuals go through when facing a major change. In the current context, this model assisted in understanding staff's emotions and resistance. Framing the turmoil as a natural part of change increased the management's willingness to persevere. Along with monitoring and staff accommodation, this prevented a premature stop of the implementation (Proctor et al., 2011).

## Conclusion

This study demonstrated that a taxonomy of outcomes from the field of implementation science provides a valuable means to evaluate the implementation of a forensic risk assessment instrument. Examining the implementation through predefined outcomes provided a wealth of information. The immediate and ongoing feedback enabled us to optimize the process. We recommend using frameworks and measures from implementation science in future implementations of risk assessment instruments. This will enhance the quality of risk assessment implementation (research), produce more solid evidence for successful implementation strategies and aid knowledge transfer. Furthermore, increasing the implementation quality of structured risk assessment instruments may in turn impact the effectiveness of these tools in reducing recidivism (Viljoen et al., 2018).

## References

- Belyh, A. (2015). *Understanding the Kübler-Ross change curve*. Retrieved January 30th, 2019, from <https://www.cleverism.com/understanding-kubler-ross-change-curve/>
- Bhanwer, A., Shaffer, C., & Viljoen, J. L. (2016). *Short-term assessment of risk and treatability: Annotated bibliography*. Burnaby, Canada: British Columbia University.
- Brown, J., & Singh, J. P. (2014). Forensic risk assessment: A beginner's guide. *Archives of Forensic Psychology, 1*, 49–59.
- Clarke, D. E., Brown, A.-M., & Griffith, P. (2010). The Brøset Violence Checklist: Clinical utility in a secure psychiatric intensive care setting. *Journal of Psychiatric and Mental Health Nursing, 17*,



- 614–620. <http://dx.doi.org/10.1111/j.1365-2850.2010.01558.x>
- Crocker, A. G., Braithwaite, E., Laferrière, D., Gagnon, D., Venegas, C., & Jenkins, T. (2011). START changing practice: Implementing a risk assessment and management tool in a civil psychiatric setting. *International Journal of Forensic Mental Health, 10*, 13–28. <http://dx.doi.org/10.1080/14999013.2011.553146>
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*, 50–66. <http://dx.doi.org/10.1186/1748-5908-4-50>
- De Beuf, T. L. F., de Ruiter, C., & de Vogel, V. (2016). *Short-term assessment of risk and treatability: Adolescenten Versie (START:AV) Handleiding*. Zetten, the Netherlands: Ottho Gerhard Helddringstichting.
- Desmarais, S. L. (2017). Commentary: Risk assessment in the age of evidence-based practice and policy. *International Journal of Forensic Mental Health, 16*, 18–22. <http://dx.doi.org/10.1080/14999013.2016.1266422>
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science, 1*, 1–3. <http://dx.doi.org/10.1186/1748-5908-1-1>
- Kroppan, E., Nasset, M. B., Nonstad, K., Pedersen, T. W., Almvik, R., & Palmstierna, T. (2011). Implementation of the Short Term Assessment of Risk and Treatability (START) in a forensic high secure unit. *International Journal of Forensic Mental Health, 10*, 7–12. <http://dx.doi.org/10.1080/14999013.2011.552368>
- Kübler-Ross, E. (1969). *On death and dying. What the dying have to teach doctors, nurses, clergy and their own families*. New York, NY: Scribner.
- Levin, S. K., Nilsen, P., Bendtsen, P., & Bulow, P. (2015). Structured risk assessment instruments: A systematic review of implementation determinants. *Psychiatry, Psychology and Law, 23*, 602–628. <http://dx.doi.org/10.1080/13218719.2015.1084661>
- Lewis, C. C., Stanick, C. F., Martinez, R. G., Weiner, B. J., Kim, M., Barwick, M., & Comtois, K. A. (2015). The Society for Implementation Research Collaboration instrument review project: A methodology to promote rigorous evaluation. *Implementation Science, 10*, 2. <http://dx.doi.org/10.1186/s13012-014-0193-x>
- Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of the Royal Society of Medicine, 104*, 510–520. <http://dx.doi.org/10.1258/jrsm.2011.110180>
- Müller-Isberner, R., Born, P., Eucker, S., & Eusterschulte, B. (2017). Implementation of evidence-based practices in forensic mental health services. In R. Roesch & A. N. Cook (Eds.), *Handbook of forensic mental health services* (pp. 443–469). London, United Kingdom: Routledge. <http://dx.doi.org/10.4324/9781315627823-16>
- Nicholls, T. L., Petersen, K. L., & Pritchard, M. M. (2016). Comparing preferences for actuarial versus structured professional judgment violence risk assessment measures across five continents: To what extent is practice keeping pace with science? In J. P. Singh, S. Bjørkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment* (pp. 127–149). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199386291.003.0008>
- Peters, D. H., Tran, N. T., & Adam, T. (2013). *Implementation research in health. A practical guide*. Geneva, Switzerland: Alliance for Health Policy and Systems Research, World Health Organization.
- Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health, 36*, 24–34. <http://dx.doi.org/10.1007/s10488-008-0197-4>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunge, A., . . . Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 65–76. <http://dx.doi.org/10.1007/s10488-010-0319-7>
- Rogers, E. M. (1962). *Diffusion of innovations*. New York, NY: Free Press of Glencoe.
- Sellers, B. G., Desmarais, S. L., & Hanger, M. W. (2017). Measurement of change in dynamic factors using the START:AV. *Journal of Forensic Psychology Research and Practice, 17*, 198–215. <http://dx.doi.org/10.1080/24732850.2017.1317560>
- Sher, M. A., & Galton, E. (2014). Implementation of the START:AV in a secure adolescent service. *Journal of Forensic Practice, 16*, 184–193. <http://dx.doi.org/10.1108/JFP-04-2013-0021>
- Singh, J. P., Desmarais, S. L., Hurdacas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., . . . Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health, 13*, 193–206. <http://dx.doi.org/10.1080/14999013.2014.922141>
- Vermaes, I., Konijn, C., Jambroes, T., & Nijhof, K. (2014). Statische en dynamische kenmerken van jeugdigen in JeugdZorgPlus: Een systematische re-



- view [Static and dynamic characteristics of youth in secured residential care: A systematic review]. *Orthopedagogiek: Onderzoek en Praktijk*, 53, 278–292.
- Viljoen, J. L., Cochrane, D. M., & Jonnson, M. R. (2018). Do risk assessment tools help manage and reduce risk of reoffending? A systematic review. *Law and Human Behavior*, 42, 181–214. <http://dx.doi.org/10.1037/%2F1hb0000280>
- Viljoen, J. L., Nicholls, T. L., Cruise, K. R., Desmarais, S. L., & Webster, C. D. (2014). *Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV): User guide*. Burnaby, British Columbia: Mental Health, Law, and Policy Institute.
- Vincent, G. M., Guy, L. S., Gershenson, B. G., & McCabe, P. (2012). Does risk assessment make a difference? Results of implementing the SAVRY in juvenile probation. *Behavioral Sciences & the Law*, 30, 384–405. <http://dx.doi.org/10.1002/bsl.2014>
- Vincent, G. M., Guy, L. S., & Grisso, T. (2012). *Risk Assessment in juvenile justice: A guidebook for implementation*. Retrieved from <http://modelsforchange.net/publications/346>
- Vojt, G., Slessor, M., Marshall, L., & Thomson, L. (2011). The clinical reality of implementing formal risk assessment and management measures within high secure forensic care. *Medicine, Science, and the Law*, 51, 220–227. <http://dx.doi.org/10.1258/msl.2011.010149>
- Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., . . . Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, 12, 108–120. <http://dx.doi.org/10.1186/s13012-017-0635-3>

Received June 18, 2018

Revision received February 6, 2019

Accepted February 26, 2019 ■

### E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <https://my.apa.org/portal/alerts/> and you will be notified by e-mail when issues of interest to you become available!