



Universiteit  
Leiden  
The Netherlands

## **Do you see what I see: observation in juvenile justice institutions**

Lampe, K.G.

### **Citation**

Lampe, K. G. (2024, March 14). *Do you see what I see: observation in juvenile justice institutions*. Retrieved from <https://hdl.handle.net/1887/3721968>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3721968>

**Note:** To cite this publication please use the final published version (if applicable).

# 5

## Risk assessment in incarcerated adolescents- Using an observation checklist to predict aggressive incidents

Kore Lampe - Maxime Kok - Rikkert van der Lans - Eva Mulder - Robert Vermeiren

*Accepted for publication in Journal of Forensic Psychology Research and Practice.*



## Abstract

Predicting risk for aggression in juvenile justice institutions (JJIs) is paramount for safety and tailoring treatment, but resources are often limited. In this study already available information and knowledge is used to predict aggression. Included were 228 juveniles aged 13–17 years. Data was used from the Reactive Proactive Aggression Questionnaire (RPQ) and the Strength and Difficulties Questionnaire (SDQ) routinely administered at intake. Knowledge from group workers was harvested by an Observation Checklist (OC) for 3–28 days. Aggressive incidents, our dependent variable, were registered for a period of 1–340 days. Poisson regression models were used to predict the number of aggressive incidents. The OC showed unique value for predicting aggressive incidents, also when controlled for the SDQ and the RPQ ( $\text{Exp}(B) = 7.483$ ,  $1.118$ ,  $p < 0.05$ ). Also the SDQ, and not the RPQ, was found to have unique value for prediction, although effects were less strong. Moreover, a predictive effect was found for the OC, the RPQ and the SDQ, on the number of incidents ( $\text{Exp}(B) = 8.197$ ,  $1.026$  and  $1.155$ , respectively). These results suggest that the OC offers unique information not incorporated in used self reports, underscoring the importance of harvesting knowledge of group workers more systematically to help identify youth at risk.

## Introduction

Aggressive incidents occur frequently in JJIs. These incidents pose a danger to the juveniles as well as to the personnel working in the JJIs. They lead to fear and stress, subsequently probably interfering with the therapeutic environment (Alink, Euser, Bakermans-Kranenburg, & van IJzendoorn, 2014). Moreover, they negatively impact job satisfaction, work performance and the well-being of personnel (Alink, Euser, Bakermans-Kranenburg, & van IJzendoorn, 2014; Harris & Leather, 2012). Risk assessment instruments developed for this population, such as the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006) require quite some information on the juvenile. Items in the SAVRY, commonly administered in Dutch JJIs after three months, ask for reflection on the behavior of the youths in the past six months and inquire about the youth's history and social context. This information is often unavailable in the first weeks or months, and sometimes remains unknown or is non-existing (for example the criminal history of first offenders). This makes instruments like the SAVRY less suited for an early taxation on risk of aggression. Thus, an improvement in the early assessment of risk of aggressive incidents is valuable for juveniles and personnel, but due to limited time and resources, this should be done preferably tapping from information and knowledge that is already collected or present.

In Dutch JJIs, a broad clinical assessment and intake procedure of detained youths is routine, mostly aimed at identifying youths that are in immediate need of care. To this end, self report inventories such as the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) are used as routine clinical assessment within the first days after entering the JJI. These instruments also contain subscales aimed at measuring externalizing behavior, e.g. the subscale Conduct Problems on the SDQ. In theory, these could possibly contribute to a risk analysis for aggressive incidents.

Although a valuable source of information (Vaughn & Howard, 2005; Butler, MacKay, & Dickens, 1995), self report inventories also bear disadvantages (Markus, Collins, Vahl, Matser, & Vermeiren, 2009). For example, due to language barriers or cultural differences in expressing certain concepts, self report can be less suited for youths with non-Dutch backgrounds (van Batenburg-Eddes et al. 2012 and Crone, Bekkema, Wiefferink & Reijneveld, 2009). A part of the juveniles, moreover, lacks the cognitive skills to adequately complete the questionnaires. Furthermore, specifically externalizing behavior has been found to be underreported by adolescents, in comparison to e.g. teachers and parents (Smith, 2007). Discrepancies between self-perceptions and other-perceptions have been described abundantly in the general

psychological literature (e.g., Kenny, Albright, Malloy & Kashy, 1994; Kenny, 2004). Hence, discrepancies are not unique to delinquent youth, although for example out of shame or out of fear that the information is revealed to a judge, these juveniles might tend to portray themselves differently by systematically overreporting socially desirable behaviors and underreporting socially undesirable behaviors (Poltavski, Van Eck, Winger, & Honts, 2018). Other-perceptions are in general considered the more reliable and accurate characterizations of other persons' behaviour (Kenny, 2004). Hence, the complementary use of self-perceptions with other-perceptions, such as parents or teachers, could help to lessen some of the above concerns but is often complex due to difficulties reaching the parents or incomplete school records (Colins et al., 2012; Kroll et al., 2002).

Observation is considered to be a relevant source of diagnostic information for various purposes, including forensic evaluation, screening and monitoring the progress of medical, psychosocial or academic interventions, and in various settings, including psychiatric hospitals, schools and prisons (Colins, Vermeiren, Schuyten, Broekaert, & Soyez, 2008; Hintze, 2005; McCann, Gilley, Hebert, Beckett, & Evans, 1997; Platzman et al., 1992; Spaans, Barendregt, Haan, Nijman, & de Beurs, 2011; Volpe, DiPerna, Hintze, & Shapiro, 2005). Moreover, research suggests that observation of aggression can be done reliably (Lampe, Mulder, Colins, & Vermeiren, 2017).

As group workers in JJIs are in regular contact with the juveniles and observe juveniles' behavior on a daily basis, researchers have suggested that structured observation by group workers in the JJIs - if carried out reliably and systematically - could prove a valuable source of information for the clinical as well as the scientific field (Hintze, 2005; McCann et al., 1997; Platzman et al., 1992; Volpe et al., 2005). Moreover, daily structured observations can possibly aid in detecting changes in state or mood, relevant for aggression (Abderhalden et al., 2008). Currently, these observations are rarely carried out structurally and systematically, and considerable variation in completeness and specificity exists, depending on factors such as group workers' characteristics, work load or amount of personnel on the shift (Lampe et al., 2017). Accordingly, although the information gathered by the group workers is valuable, it cannot be readily used in clinical practice. The clinical challenge is, therefore, to guide the group workers in carrying out these observations structurally and systematically, so it becomes possible to utilise the valuable knowledge that is already present.

The Observation Checklist (OC) guides daily observation of behavior and mood of the juveniles on six concepts: proactive aggression, reactive aggression, hyperactivity, impulsivity, signs of depressed mood and lack of reciprocity in contact. The OC

was implemented in two Dutch JJIs and was developed to complement diagnostic information, guide interventions and aid (e.g. risk management, treatment) decision making by focusing on a wide array of key concepts.

This study uses the aggression concepts from the OC, and investigates whether systematically and structurally guiding the daily observations of group workers in JJIs can aid the identification of juveniles at a higher risk of causing aggressive incidents. Moreover, we are aiming to test the same for the self report measures from the intake routine. Thus, we first examine the predictive validity of the Observation Checklist and of the self report questionnaires on aggressive incidents in the JJI. Second, the incremental validity of the OC is analysed when added to the existing self report questionnaires in order to determine the added value of this new observation method.

## Materials and methods

### Context of this study

The Netherlands count approximately 500 juveniles who are detained in juvenile justice institutions (JJIs) daily (Dit is DJI, 2019; Grisso, Vincent, & Seagrave, 2005). A large majority (>85%) of the juveniles placed in these JJIs are in pre-trial detention and have been accused of offences ranging from minor crimes, such as property or drug offences, to major crimes such as burglary or (attempted) murder. Ethnic minorities are overrepresented in the Dutch justice system and JJIs (Boon, Dorp, & Boer, 2018). Furthermore, high rates of mental disorders (Colins et al., 2010; Teplin, Abram, McClelland, Dulcan, & Mericle, 2002) and serious conduct problems (Colins et al., 2008) characterize this complex group. When placing youths in JJI, the Dutch juvenile justice system aims to maintain public safety (1), to safeguard their development (e.g. by addressing treatment needs (2)) and to reintegrate youth into the community, without re-offending (3). The second and third of these goals assign JJIs a therapeutic (and pedagogical) function. A central condition to achieve this therapeutic function is the establishment of a safe environment.

The OC is a new assessment tool developed by the Academic Workplace Forensic Care for Youth (AWFCY). The AWFCY is a Dutch collaboration between two JJIs, two universities, two child- and adolescent psychiatry centres and two colleges striving to attend to the needs of juveniles in detention, the public safety and prevention of criminal recidivism. By combining practice, research, and education clinicians, researchers, educators and policy makers are provided an opportunity to learn from each other and to enrich the knowledge and skills of employees. Another important

objective is to develop methods or instruments that can be used in practice and education, such as the Observation Checklist (OC). The OC guides daily observation of behavior and has been implemented in two Dutch JJIs. A group of twelve group workers were extensively trained for two full days by using video material. A training session of one full day followed three months after implementation. After this follow up session, reliability was researched using video excerpts, and turned out to be fair for the concept of Aggression (Lampe, Mulder, Vermeiren, & Colins, 2023). Booster sessions of three hours were regularly organised afterwards. Group workers were instructed to complete the OC at the end of their shift together with their colleague, to avoid missing information, which has resulted in adequate completion rates. Observation took place in the hours that juveniles resided at the group, for example having dinner or playing table tennis, or when they participated in activities guided by their groupworkers, such as sports or family visits. In total, observation took place approximately nine hours per day on weekdays, and eleven hours in the weekend. Steps to undertake for a facility seeking to adopt the OC approach, would be manageable- roughly taking a 3-day investment for training and follow up, and an organising body to follow up the process. An instruction manual and a training program, including video material and app are available on request. Also, more information on the development and implementation is described in detail elsewhere (Lampe, Mulder, Vermeiren, & Colins, 2023).

### Study sample

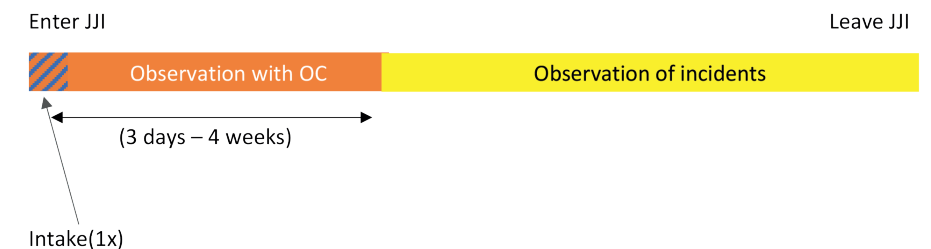
All 435 juveniles admitted to the participating Dutch JJI between February 1st 2013 and October 1st 2014 were observed using the OC and their aggressive incidents were subsequently registered until March 2016. Exclusion criteria for this study include less than five observations of the juvenile (64 juveniles), failure to complete the intake procedure (114 juveniles), and an age of 18 years or older at entry (29 juveniles). Hence, our final sample comprised 228 juveniles (see Table 1). An overrepresentation of ethnic minorities was found in our sample as can be seen in Table 1. For ethnic categorisation, we used the definition provided by the Dutch Central Bureau of Statistics (2012).

### Procedure

Data collection followed a three-stage procedure. Figure 1 visualizes the stages in colours. The blue orange represents the intake routine which took place within the first 24 hours after admission. The intake routine comprised a standard mental health screening using various validated self report questionnaires. Orange represents the stage at which juveniles were observed with the OC. Observations were performed by the group workers of the influx group during their shifts, twice daily on weekdays and once daily during the weekend. Group workers received concise instructions how to use the OC to ensure consistent use. The time of observations spanned minimally three

days (five observations) to approximately four weeks (48 observations). After two to four weeks, juveniles were placed in another group with different group workers who did not use the OC, ending the OC observation period. Grey represents the final stage at which aggressive incidents were registered. This stage spanned the remaining time that the juveniles were in the JJI, in our sample this was maximally 340 days.

**Figure 1** The timeline of the procedure of data collection in the JJI.



### Measures

#### *Intake: Strength and Difficulties Questionnaire (SDQ)*

The SDQ was designed for psychosocial assessment of children and adolescents between four and sixteen years (Goodman, 1997), consisting of five subscales of five items with three answer options: 0 = Not True, 1 = Somewhat True and 3 = Certainly True. Subscales have total scores between 0 and 10. In this study only the subscale Conduct Problems of the Dutch self report version was used, of which an alpha of .47 was found in earlier studies (van Widenfelt, Goedhart, Treffers, & Goodman, 2003). Example items were: 'Often has temper tantrums or hot tempers' and 'Often fights with other children or bullies them'. The authors that developed and researched the Dutch version, mention as a possible explanation for this low alpha, the possibility that youths only partly self report on conduct problems (van Widenfelt, Goedhart, Treffers, & Goodman, 2003), which is in line with earlier findings (Smith, 2007).

#### *Intake: Reactive Proactive Aggression Questionnaire (RPQ)*

The RPQ was used to operationalize juveniles' self-perception of their reactive aggression and their proactive aggression (Raine et al., 2006). Items each had three response options: 0 = Never, 1 = Sometimes and 2 = Often. Total scores for the two subscales range between 0 and 24 and 0 and 22 for RA and PA respectively. Prior research supports the validity of the RPQ from Dutch kids from six years old (Cima, Raine, Meesters, & Popma, 2013). The Reactive Aggression (RA) subscale consisted of

12 items ( $\alpha=0.85$ ). An example item was: 'Yelled when annoyed'. The PA subscale consisted of 11 items ( $\alpha=0.86$ ). An example item was 'Threatens and bullies'.

#### *Observation: the Observation Checklist (OC)*

The OC has group workers score six concepts holistically on a three-point Likert scale each shift, e.g. on weekdays twice a day, in weekends once a day. The six concepts are proactive aggression, reactive aggression, hyperactivity, impulsivity, signs of depressed mood and lack of reciprocity in contact. Only the two aggression scales are used in this study. The three-point scale includes the categories: 0= None, 1= Some and 2 = Much. The available number of ratings varied according to the length of incarceration. For both subscales the mean score over all observations served as the indicator of reactive and proactive aggression. The total scores were computed as the mean score of the reactive and proactive aggression scores.

#### *Follow-up: Youth Information System (YIS) and Incident Classification Manual*

All incidents and corresponding disciplinary measures in the JJI were routinely registered in a digital incident registration system, the Youth Information System (YIS). Incidents that succeeded the last observation with the OC for that specific juvenile, and were registered between January 2013 and February 2016, were used. We categorized the incidents into aggression (e.g. physical and verbal aggression), opposition/ mild aggression (e.g. oppositional, disruptive, or disrespectful behavior), withdrawal-related, substance-related, and other by using a manual developed for this purpose. Interrater reliability of the categorization was high (percentage of agreement of 95%) and was tested by having a colleague classifying a random sample of 20 incidents. For this study, only the incidents classified as aggressive were used.

### **Data handling and statistical analysis**

Data handling and statistical analyses were performed using Microsoft Excel and IBM SPSS Statistics 25. The dependent variable (number of aggressive incidents) is considered to be count variable ranging between 0 (no incidents reported) and 14. Independent variables included in the models were OC's Proactive, Reactive and Total aggression, RPQ Total aggression, and SDQ Conduct Problems.

Because count variables violate assumptions behind regular regression models (Gardner, Mulvey, & Shaw, 1995), Poisson regression models were used to assess predictive and incremental validity of the OC. Various other negative binominal regression models were estimated but the fit was less.

All Poisson regression models were estimated in SPSS using non-linear generalized linear models in combination with an exponential link function. An offset variable was included in the models to adjust estimations for the varying length of the follow-up period. The offset quantified the number of days that incidents were registered (between 1 - 340 days). Fit of the Poisson model was assessed by the deviance statistic and by evaluation of under- and overdispersion using the  $\chi^2/df$  ratio statistic. Criteria for the  $\chi^2/df$  were based on Payne, Gebregziabher, Hardin, Ramakrishnan and Egede (2018), that Poisson regression models  $\chi^2/df > 1.2$  suggest overdispersion. Reported significance tests applied an alpha level of 0.05.

Predictive validity of the OC, the RPQ and SDQ was examined using Poisson models that predicted the Number of Aggressive Incidents by one of the following independent variables: OC Total aggression, OC, Proactive aggression, OC Reactive aggression, RPQ Total aggression, RPQ Proactive Aggression, RPQ Reactive Aggression and SDQ Conduct Problems. We note that the different ranges of the independent variables need to be considered when interpreting the  $\exp(B)$  effects, e.g. the OC has scores under 1 whilst the RPQ ranges from 0-18.

To address our question on the incremental validity of the OC, a hierarchical regression procedure was performed. These Poisson regression models estimated the unique effect of the last entered independent variable controlled for the effects of the other independent variables. We performed this analysis first in three different orders: with all total aggression scales (e.g. OC Total aggression, RPQ Total aggression and SDQ Conduct Problems) each in turn last. Then we explored the different aggression types, by using the specific RA or PA scales of the OC and the RPQ, with per type three analyses with each instrument entered in turn last (for example SDQ, RPQ RA and OC RA in one analysis).

To test whether the OC Total aggression, OC Proactive aggression, OC Reactive aggression, RPQ Total aggression, RPQ Proactive Aggression, RPQ Reactive Aggression and SDQ Conduct Problems separately predict Number of Aggressive Incidents in the JJI, another set of Poisson analyses was performed. Models included the dependent variable, the offset variable and, in turn, one of the independent variables.

## **Results**

### **Descriptive statistics**

Table 1 shows the descriptive statistics of the data. The outcome variable Number of Aggressive incidents ranged between 0 to 14, but for most juveniles it was 0, which is reflected by the median (see Table 1).

**Table 1** Study sample characteristics and descriptive statistics

	<b>n</b>	<b>%</b>	<b>M</b>	<b>SD</b>	<b>Range</b>
Age (Years)	228		16.42	1.13	13.12-17.98
Ethnicity					
Dutch	25	11.0			
Moroccan	69	30.3			
Turkish	18	7.9			
Surinamese	25	11.0			
Dutch-Antillean	21	9.2			
Other	70	30.7			
Duration of stay in JJI (Days)			79.90	113.45	3-1177
Incident Registration Period (Days)			56.69	82.04	1-340
OC Total Aggression			0.11	0.16	0-0.91
OC Proactive Aggression			0.03	0.07	0-0.50
OC Reactive Aggression			0.08	0.12	0-0.72
RPQ Total Aggression			8.63	6.47	0-32
RPQ Proactive Aggression			2.12	2.87	0-14
RPQ Reactive Aggression			6.66	4.23	0-18
SDQ Conduct Problems			1.74	1.50	0-9
Number of Aggressive incidents			1.23	2.30	0-14

Note. JJI: Juvenile Justice Institution, OC: Observation Checklist; RPQ: Reactive Proactive Aggression Questionnaire, SDQ: Strengths and Difficulties Questionnaire.

### Incremental and predictive validity

Results of both sets of analyses can be found in Table 2. A significant Wald Chi-Square statistic was found for the OC total aggression score controlled for the RPQ and SDQ total scores. Significant Wald Chi-Square statistic were also found for the OC RA and OC PA subscales when controlled for the RPQ RA and PA subscales and SDQ conduct subscale score. With respect to the self-perception questionnaires, we found a significant Wald Chi-Square statistic for the SDQ Conduct Problems, but no significant effect of the RPQ - not for RA subscale, PA subscale, or their total score - controlling for the other independent variables.

We found evidence of predictive validity for all independent variables (see Table 2). The Exp(B), also known as the incidence rate ratio or relative risk of getting the outcome, indicates that all independent variable were predictive of the observed

Number of Aggressive Incidents. These results corroborates the utility of the self report instruments and the OC for predicting aggressive incidents.

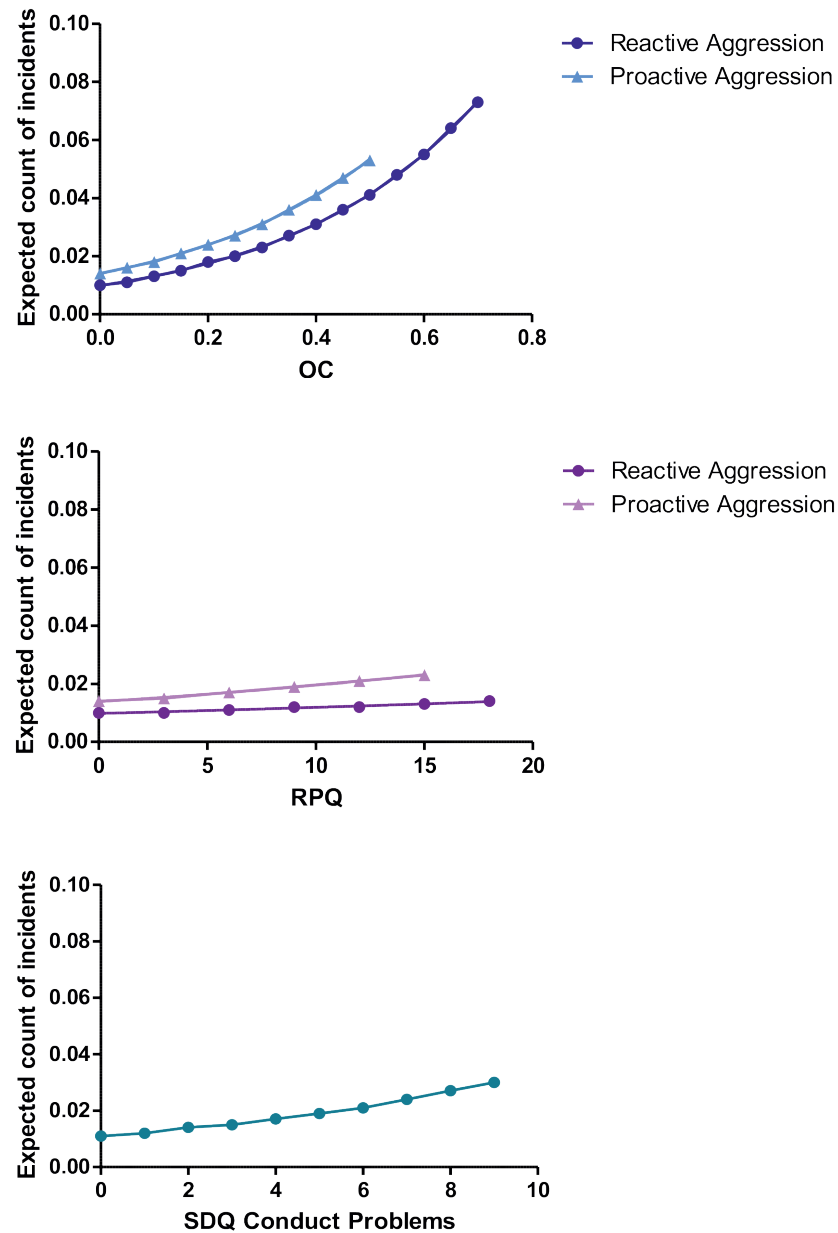
**Table 2** Hierarchical and single Poisson analyses

		$\chi^2/df$	Exp(B)	CI for EXP (B) 95%	
				Lower	Upper
<b>Incremental validity (hierarchical Poisson)</b>					
<i>Var. entered</i>	<i>Var. entered last</i>				
RPQ TA SDQ OC TA	OC TA	1.28	7.48*	4.319	12.967
	SDQ	1.28	1.12*	1.024	1.221
	RPQ TA	1.28	1.01	0.990	1.032
RPQ RA SDQ OC RA	OC RA	1.29	17.35*	8.489	35.450
	SDQ	1.29	1.14*	1.047	1.245
	RPQ RA	1.29	1.02	0.986	1.049
RPQ PA SDQ OC PA	OC PA	1.45	14.56*	3.099	68.434
	SDQ	1.45	1.11*	1.018	1.204
	RPQ PA	1.45	1.03	0.985	1.086
<b>Predictive validity (single independent predictor Poisson)</b>					
	OC TA	1.31	8.20*	4.669	14.391
	OC RA	1.32	16.28*	7.871	33.690
	OC PA	1.44	20.546*	4.737	89.120
	RPQ TA	1.61	1.026*	1.008	1.045
	RPQ RA	1.61	1.029*	1.002	1.057
	RPQ PA	1.60	1.078*	1.035	1.124
	SDQ	1.54	1.155*	1.074	1.242

Note. Var. = variable, RPQ= Reactive Proactive Aggression Questionnaire, SDQ= Strengths and Difficulties Questionnaire (conduct disorder scale), OC= Observation Checklist, PA= Proactive aggression, RA= Reactive aggression, TA=Total aggression, CI = confidence interval.

The slope of the graphs of the models in Figure 2 illustrate the incremental effect of each independent variable when controlled for the effects of the other two measures. For clarity purposes, only the aggression subtypes and not the total scores are presented, as ranges/scales differ on the X-axis.

**Figure 2** Hierarchical Poisson Regression models with number of aggressive incidents and incremental variables per aggression type, controlled for the other measures.



*Note.* The range of the x-axis was based on the range of values of the predictor variables.

Differences in  $\exp(B)$  do not directly reflect the relevance of the different independent variables to predict risk on aggressive incidents. This is because the independent variables have different range, for example, the highest score of the OC is <1 and the highest score of the RPQ total aggression is 32 in our sample. Thus, interpretations of this  $\exp(B)$  need to take this in account: a step of '1' on the X axis, covers the whole range when looking at the OC, yet is only a small step on the RPQ. For illustration purposes, we estimated expected counts of incidents using the Poisson models for some fictive examples.

*Scenario 1 (RPQ)*

Suppose that juvenile A scores 9 on the RPQ (the median score in this sample) and suppose that juvenile B would score 32 (highest score), then the chance that group workers experience an incident with juvenile B is estimated to be 1.85 times higher compared to change of experiencing an incident with juvenile A.

*Scenario 2 (SDQ)*

Suppose that juvenile A scores 2 on the SDQ (the median score in this sample) and suppose that juvenile B scores 9 (the highest score), then the chance that group workers experience an incident with juvenile B is estimated to be 2.80 times higher compared to change of experiencing an incident with juvenile A.

*Scenario 3 (OC Checklist)*

When juvenile A would be observed with the OC for 10 shifts, and only once is scored a 1 on the checklist - implying a mean OC score of 0.1 (rather typical for the current sample). Juvenile B, instead, was scored a total score of 9 in 10 shifts (the highest number within 10 shifts in our sample), then the chance that group workers experience an incident with juvenile B is estimated to be 5.40 times higher than that of juvenile A.

**Discussion**

Our results provide evidence that observing aggression using the OC is of unique value in identifying youths at higher risk for aggressive incidents in JJIs. Both proactive, reactive and total aggression, when observed with the OC, have an added value over the self report measures in the prediction of the number of aggressive incidents. Also the conduct problems scale of the SDQ adds unique value in predicting aggressive incidents, albeit less strong. No unique effects of the RPQ scales were found when added last, meaning no contribution to the prediction is found when the OC and the SDQ are already administered. Moreover, both aggression scores of the OC, both

aggression scores of the RPQ and the conduct scale of the SDQ were all separately predictive of aggressive incidents.

Our results suggest that the OC scale is the strongest predictor of future aggressive incidents, since the exponentiated regression coefficient of the Poisson Regression,  $\text{Exp}(B)$ , can be interpreted as the incidence rate ratio or relative risk of getting the outcome. The fact that previous (observed) aggression predicts future aggression, is not new and well-known (Farrington, 1991; Tolan, Guerra, & Kendall, 1995). Other instruments, such as the Brøset Violence Checklist (Woods & Almvik, 2002) and the Dynamic Appraisal of Situational Aggression (DASA) (Ogloff and Daffern, 2006), show promising results also in youth (Dickens, O'Shea, & Christensen, 2020). The BVC and DASA approach overlaps with the OC in that it also relies on group workers knowledge, but have a different angle as they are developed for a psychiatric population and are targeting imminent risk of aggression. Also, the BVC includes a more dynamic approach, e.g. confusion is only scored when a patient is more confused than 'normally'- rendering it suitable for patients already known by the observing nurses. Research on these instruments underscores, however, that a systematic approach in observing and administering (signs of) aggression is very suitable in assessing risk (Chu, Daffern & Ogloff, 2013). To our knowledge, the OC is the only broader looking observation checklist of which the structured observations of aggression from within the JJI were used to predict aggression in the same setting later on, and the results are promising.

Although researchers regularly point to the possibility of using observations in JJIs by group as a valuable source of clinical and scientific information, these observations are rarely carried out systematically (Hintze, 2005; Lampe et al., 2017; McCann et al., 1997; Platzman et al., 1992; Volpe et al., 2005), leaving important information unseen and unreported. The OC bridges this gap by making this often implicit knowledge, explicit and reporting it systematically. The gained competence in group workers is already a win, while this study shows also the value of using this long-hidden information for the identification of youths at higher risk of aggressive incidents, by using observation. While the current study only uses the aggression concepts of the OC, deliberately, the OC focuses on a wider array of relevant constructs for diagnostics and treatment planning. This is important because it broadens the view of groupworkers and forces to also focus on more internalizing or other behavior, leading to, first, an educational and empowering profit and second, to more information for assessment.

Identification of youths at risk for causing incidents in the JJI is relevant in many ways. First, as time and resources in JJIs are generally limited, and more extensive assessment

can also be burdening for youths, the identification of youths at higher risk could be used for triage purposes for assignment to more in-depth assessment (e.g. clinical interviews in addition to self reports). Second, it can help guide treatment planning, e.g. type, form and intensity of treatment. As mental health disorders are highly prevalent in JJIs (Abram, Teplin, McClelland, & Dulcan, 2003), an effective diversion to treatment programmes is needed and tailoring these to the juveniles needs and strengths crucial (Hillege, Brand, Mulder, Vermeiren, & van Domburgh, 2017). Third, it can warrant a closer monitoring for safety concerns, safeguarding both the juveniles health and that of personnel. Examples could be placing youths who were observed showing more aggression in smaller groups, spreading them through the institution, or assigning additional qualified personnel- in this way hopefully reducing aggression. Fourth, the structural collection of these observations can possibly aid the clinician in performing risk assessment. In these ways, the knowledge of the group workers in JJIs can ultimately be utilised for the improvement of the juveniles' therapeutic care and the safety in JJIs.

Notwithstanding the promising results of the OC, we must emphasize that we should be prudent at all times to not solely rely on a single source of information. In line with other scholars (Colins et al., 2008), we stress the importance of keeping a broad view of the juvenile and its system, and of the involvement of as many sources of information as possible, e.g. parents, teacher, self report and historical reports, as this can contribute to a comprehensive and conscientious assessment that benefits tailored treatment.

A methodological aspect of this study that calls for further review is the assumption of the Poisson Distribution that the observations of the outcome are independent. One could argue that due to interactions between juveniles, group workers, their perceptions and experience with each other, and group climate, this independency is not entirely met. However, Poisson regression has been previously used for similar data (e.g. incidences of domestic violence) (Famoye & Singh, 2006) and deemed fit.

This study gives rise to a wide array of suggestions for further research. While the predictive effect of certain subscales and composite scores of the OC and the self report measures on the number of aggressive incidents was investigated in this study, other scales are yet left unconsidered. It would be interesting to analyse the predictive validity of the hyperactivity and impulsivity scales of the OC and of corresponding self report scales, since impulsive aggression and hyperactivity frequently co-occur in youth with Attention-Deficit/Hyperactivity Disorder (Connor, Chartier, Preen, & Kaplan, 2010; Saylor & Amann, 2016). Also, as intellectual disabilities can influence

validity of self report (Finlay & Lyons, 2001), one could study whether self report is less predictive in a cognitively challenged group. Another suggestion for further research is to investigate the role of ethnicity in the predictive validity of the OC and the two self report measures. In concordance with other findings in Dutch JJIs, our study sample showed an overrepresentation of ethnic minorities (Komen & Van Schooten, 2009; van Batenburg-Eddes et al., 2012). Moreover, some minorities have found to report less externalizing problems on self reports (Crone, Bekkema, Wiefferink, & Reijneveld, 2010; van Batenburg-Eddes et al., 2012). Adding OC data to this complex matter proved to be able to help shedding some more light on this multifaceted challenge- see also Lampe, Mulder and Vermeiren (2023).

The Observation Checklist was developed in a multidisciplinary effort meant to bridge the gap between practice, science and policy. Our findings in this study prove the development and implementation exceed this expectation by its usefulness in predicting aggression in the JJI. Our results are uplifting in a way that the SDQ and the RPQ, currently already part of the standard intake routine in JJIs, both can aid in the prediction aggressive incidents. Promisingly, our data also show that the OC was a stronger predictor of future aggressive incidents than the self report scales. Consequently we urge policymakers and JJIs to think of a way to structurally collect these observations, for example by using the (free available) OC, and use them for risk assessment and treatment planning. Moreover, as the self report data also showed predictive qualities, we would wish them also to develop a way on how to make use of the self report data more systematically as to involve them in triage and forecasting purposes.

## References

- Abderhalden, Christoph, et al. "Structured risk assessment and violence in acute psychiatric wards: randomised controlled trial." *The British Journal of Psychiatry* 193.1 (2008): 44-50.
- Abram, K. M., Teplin, L. A., McClelland, G. M., & Dulcan, M. K. (2003). Comorbid psychiatric disorders in youth in juvenile detention. *Arch Gen Psychiatry*, 60(11), 1097-1108. doi:10.1001/archpsyc.60.11.1097
- Alink, L. R. A., Euser, S., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2014). A challenging job: Physical and sexual violence towards group workers in youth residential care. *Child & Youth Care Forum*, 43(2), 243-250. doi:10.1007/s10566-013-9236-8
- Boon, A., Dorp, M., & Boer, S. (2018). Oververtegenwoordiging van jongeren met een migratieachtergrond in de strafrechtketen. *Tijdschrift voor Criminologie*, 3, 268-288.
- Borum, R., Bartel, P. & Forth, A. (2002). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)*. Tampa: University of South Florida.
- Butler, S. M., MacKay, S. A., & Dickens, S. E. (1995). Maternal and adolescent ratings of psychopathology in young offender and non-clinical males. *Canadian Journal of Behavioral Science/Revue Canadienne des Sciences du Comportement*, 27, 333.
- Chi Meng Chu, Michael Daffern & James R.P. Ogloff (2013) Predicting aggression in acute inpatient psychiatric setting using BVC, DASA, and HCR-20 Clinical scale, *The Journal of Forensic Psychiatry & Psychology*, 24:2, 269-285, DOI: 10.1080/14789949.2013.773456
- Colins, O., Vermeiren, R., Schuyten, G., Broekaert, E., & Soyez, V. (2008). Informant agreement in the assessment of disruptive behavior disorders in detained minors in Belgium: A diagnosis-level and symptom-level examination. *The Journal of Clinical Psychiatry*, 69(1), 141-148. doi:10.4088/JCP.v69n0119
- Colins, O., Vermeiren, R., Vahl, P., Markus, M., Broekaert, E., & Doreleijers, T. (2012). Parent-reported attention-deficit hyperactivity disorder and subtypes of conduct disorder as risk factor of recidivism in detained male adolescents. *European Psychiatry*, 27(5), 329-334. doi:10.1016/j.eurpsy.2011.01.001
- Colins, O., Vermeiren, R., Vreugdenhil, C., van den Brink, W., Doreleijers, T., & Broekaert, E. (2010). Psychiatric disorders in detained male adolescents: a systematic literature review. *Can J Psychiatry*, 55(4), 255-263. doi:10.1177/070674371005500409
- Connor, D. F., Chartier, K. G., Preen, E. C., & Kaplan, R. F. (2010). Impulsive aggression in attention-deficit/hyperactivity disorder: symptom severity, co-morbidity, and attention-deficit/hyperactivity disorder subtype. *Journal of child and adolescent psychopharmacology*, 20(2), 119-126. doi:10.1089/cap.2009.0076
- Crone, M. R., Bekkema, N., Wiefferink, C. H., & Reijneveld, S. A. (2010). Professional identification of psychosocial problems among children from ethnic minority groups: room for improvement. *The Journal of pediatrics*, 156(2), 277-284.

- Dickens, G. L., O'Shea, L. E., & Christensen, M. (2020). Structured assessments for imminent aggression in mental health and correctional settings: Systematic review and meta-analysis. *International journal of nursing studies*, 104, 103526.
- Dit is DJI. (2019, January). Retrieved from [https://www.dji.nl/binaries/Infographic\\_DitIsDJI\\_V2\\_tcm41-286791.pdf](https://www.dji.nl/binaries/Infographic_DitIsDJI_V2_tcm41-286791.pdf)
- Famoye, F., & Singh, K. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, 4, 117-130.
- Farrington, D. P. (1991). Childhood aggression and adult violence: Early precursors and later-life outcomes. In *The development and treatment of childhood aggression*. (pp. 5-29). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Finlay, W., & Lyons, E. (2001). Methodological issues in interviewing and using self report questionnaires with mental retardation. *Psychological assessment*, 13, 319-335. doi:10.1037/1040-3590.13.3.319
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol Bull*, 118(3), 392-404. doi:10.1037/0033-2909.118.3.392
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Child Psychology & Psychiatry & Allied Disciplines*, 38(5), 581-586. doi:10.1111/j.1469-7610.1997.tb01545.x
- Grisso, T., & Barnum, R. (2000). Massachusetts youth screening instrument second version. Users manual and technical report.
- Grisso, T., Vincent, G., & Seagrave, D. (2005). Risk for Violence and Recidivism Assessment Tools. In *Mental Health Screening and Assessment in Juvenile Justice* (pp. 265-268). New York: The Guilford Press.
- Harris, B., & Leather, P. (2012). Levels and consequences of exposure to service user violence: Evidence from a sample of UK social care staff. *British Journal of Social Work*, 42(5), 851-869. doi:10.1093/bjsw/bcr128
- Hillege, S. L., Brand, E., Mulder, E. A., Vermeiren, R., & van Domburgh, L. (2017). Serious juvenile offenders: classification into subgroups based on static and dynamic characteristics. *Child Adolesc Psychiatry Ment Health*, 11, 67. doi:10.1186/s13034-017-0201-4
- Hintze, J. M. (2005). Psychometrics of Direct Observation. *School Psychology Review*, 34(4), 507-519. Retrieved from <https://login.ezproxy.leidenuniv.nl:2443/login?URL=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2006-02180-005&site=ehost-live> hintze@educ.umass.edu
- Komen, M., & Van Schooten, E. (2009). Ethnic Disparities in Dutch Juvenile Justice. *Journal of Ethnicity in Criminal Justice*, 7(2), 85-106. doi:10.1080/15377930902929182
- Kroll, L., Rothwell, J., Bradley, D., Bailey, S., Shah, P., & Harrington, R. C. (2002). Mental health needs of boys in secure care for serious or persistent offending: A prospective, longitudinal study. *The Lancet*, 359(9322), 1975-1979. doi:10.1016/S0140-6736(02)08829-3
- Lampe, K. G., Mulder, E. A., Colins, O. F., & Vermeiren, R. R. J. M. (2017). The inter-rater reliability of observing aggression: A systematic literature review. *Aggression and Violent Behavior*, 37, 12-25. doi:10.1016/j.avb.2017.08.001
- Lampe, K. G., Mulder, E. A., & Vermeiren, R. R. (2023). Ethnic differences in assessment: How self report and observation converge and diverge among ethnically diverse incarcerated youths. *Journal of Threat Assessment and Management*.
- Lampe, K. G., Mulder, E. A., Vermeiren, R. R., & Colins, O. F. (2023). An observation checklist for use by residential social workers in juvenile justice institutions. *Journal of Social Work*, 14680173231164291.
- McCann, J. J., Gilley, D. W., Hebert, L. E., Beckett, L. A., & Evans, D. A. (1997). Concordance between direct observation and staff rating of behavior in nursing home residents with Alzheimer's disease. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, 52(2), P63-P72. doi:10.1093/geronb/52B.2.P63
- Payne, E. H., Gebregziabher, M., Hardin, J. W., Ramakrishnan, V., & Egede, L. E. (2018). An empirical approach to determine a threshold for assessing overdispersion in Poisson and negative binomial models for count data. *Commun Stat Simul Comput*, 47(6), 1722-1738. doi:10.1080/03610918.2017.1323223
- Platzman, K. A., Stoy, M. R., Brown, R. T., Coles, C. D., Smith, I. E., & Falek, A. (1992). Review of observational methods in attention deficit hyperactivity disorder (ADHD): Implications for diagnosis. *School Psychology Quarterly*, 7(3), 155-177. doi:10.1037/h0088258
- Poltavski, D., Van Eck, R., Winger, A. T., & Honts, C. (2018). Using a Polygraph System for Evaluation of the Social Desirability Response Bias in Self report Measures of Aggression. *Appl Psychophysiol Biofeedback*, 43(4), 309-318. doi:10.1007/s10484-018-9414-4
- Saylor, K. E., & Amann, B. H. (2016). Impulsive Aggression as a Comorbidity of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Journal of child and adolescent psychopharmacology*, 26(1), 19-25. doi:10.1089/cap.2015.0126
- Smith, S. R. (2007). Making Sense of Multiple Informants in Child and Adolescent Psychopathology: A Guide for Clinicians. *Journal of Psychoeducational Assessment*, 25(2), 139-149. doi:10.1177/0734282906296233
- Spaans, M., Barendregt, M., Haan, B., Nijman, H., & de Beurs, E. (2011). Diagnosis of antisocial personality disorder and criminal responsibility. *International Journal of Law and Psychiatry*, 34(5), 374-378. doi:10.1016/j.ijlp.2011.08.008
- Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in juvenile detention. *Arch Gen Psychiatry*, 59(12), 1133-1143. doi:10.1001/archpsyc.59.12.1133
- Tolan, P. H., Guerra, N. G., & Kendall, P. C. (1995). A developmental-ecological perspective on antisocial behavior in children and adolescents: toward a unified risk and intervention framework. *J Consult Clin Psychol*, 63(4), 579-584. doi:10.1037//0022-006X.63.4.579
- van Batenburg-Eddes, T., Butte, D., van de Looij-Jansen, P., Schiethart, W., Raat, H., de Waart, F., & Jansen, W. (2012). Measuring juvenile delinquency: How do self reports compare with official police statistics? *European Journal of Criminology*, 9(1), 23-37. doi:10.1177/1477370811421644
- Vaughn, M. G. & Howard, M. O. (2005). Self report measures of juvenile psychopathic personality traits: A comparative review. *Journal of Emotional and Behavioral Disorders*, 13, 152-162.

- van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry, 12*(6), 281-289. doi:10.1007/s00787-003-0341-3
- Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing Students in Classroom Settings: A Review of Seven Coding Schemes. *School Psychology Review, 34*(4), 454-474.
- Woods, P., & Almvik, R. (2002). The Brøset violence checklist (BVC). *Acta Psychiatrica Scandinavica, 106*, 103-105.