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

Lampe, K.G.

4

Cultural Factors in Risk Assessment

Kore Lampe - Eva Mulder - Robert Vermeiren

Published in Journal of Threat Assessment and Management, 10(2), 112–125 (2023).

Abstract

While structured observation and self report are of great value for risk assessment in forensic youth psychiatry, both methods are culturally sensitive. This study therefore researches data harvested by self report using standardized questionnaires, and by an Observation Checklist (OC) in an ethnically diverse population of incarcerated youths. Our population consisted of 228 incarcerated male juveniles, with the majority, 30.2%, of Moroccan descent, 11.2% of Dutch origin, 11.2% of Surinamese descent, 9.1% with a Netherlands Antilles background and 8.2% of Turkish origin. Other ethnicities, or adolescents of whom ethnicity data were missing, constituted 30.1% percent of the final sample. First, scores on each self report subscale and OC concept were analysed for differences between ethnic backgrounds, second, OC concepts were matched to relevant self report scales, e.g. proactive aggression of the OC to proactive aggression of the RPQ, and finally convergence and divergence between the two matched concepts was analysed. Large differences were found for the separate methods, and the divergence and convergence between the two methods. Most prominently a very different scoring profile between youths of Dutch and Moroccan backgrounds was found, with the latter self reporting fewer problems than youths of Dutch origin, while more problems were observed. Possible explanations, such as (racial) bias in observing, lack of cross cultural validation for self report, or biaises such as shame, fear of judicial consequences are discussed.

Introduction

The assessment and evaluation of incarcerated juveniles is of key importance for both risk assessment and tailored treatment. However, assessment of these juveniles remains a complex challenge due to several reasons. Reliability and validity of assessment vary as a function of different factors, such as the role and age of the informant, the type of behavior, features of the setting and the type of interviewer and instrument (Achenbach, McConaughy, & Howell, 1987; Ferdinand, van der Ende, & Verhulst, 2004; Florsheim, Shotorbani, Guest-Warnick, Barratt & Hwang, 2000; Fazel, Doll & Longstrom, 2008). Different methods and sources of information (e.g. parents, youths, teachers, observations by group workers) often do not correlate (Janssen, Verhulst, Bengi-Arslan, Erol, Salter, & Crijnen, 2005; Colins, Vermeiren, Schuyten, Broekaert, & Soyez, 2008; Youngstrom, Loeber, & Stouthamer-Loeber, 2000; Smith, 2007).

Another important factor to consider is the role cultural factors play in assessment. Ethnic minorities are overrepresented in the judiciary system (Bishop and Frazier, 1996; Veen, Stevens, Andershed, Raaijmakers, Doreleijers, & Vollebergh, 2011), while assessment methods do not always fit this diverse population. For example, self report questionnaires can be less suited for ethnic minorities. This can be due to a lack of conceptual equivalence, the questionnaires not being cross culturally validated, language difficulties or barriers in the expression and identification of problems (van Batenburg-Eddes et al. 2012; Crone, Bekkema, Wiefferink & Reijneveld, 2009; Sue, 1995). Also, research shows that detained youths are likely unreliable in (self) reporting impairments, due to biases like social desirability, shame, insufficient introspection or cognitive delays (Ladd & Kochenderfer-Ladd, 2002; van Widenfelt et al. 2003), factors that can partly be influenced by cultural customs and norms. Not only self report, also a structured professional judgment tool such as the Structured Assessment of Violent Risk in Youth (SAVRY), has been found to be of less predictive value in different ethnic groups (Shepherd, Luebbers, Ferguson, Ogloff & Dolan, 2014; Muir, Viljoen, Jonnson, Cochrane, & Rogers, 2020).

While ethnic minorities are overrepresented in the judiciary system, they are underrepresented in the Dutch mental health care system (De Haan, Boon, Vermeiren, & De Jong 2012). When considering that prevalence of psychiatric disorders in this group is found to be similar to Dutch youths, it is suggested that psychiatric problems are addressed too late, and only after these juveniles end up in the judiciary system, instead of an earlier intervention leading them to regular mental health care (Adriaanse, van

Domburgh, Veling, & Doreleijers, 2011). This underscores the importance of timely, and above all, cultural sensitive assessment.

Psychiatric disorders elevate the risk for various detrimental outcomes, including recidivism (e.g. Fite, Raine, Stouthamer-Loeber, Loeber, & Pardini, 2009; Colins, Van Damme, Fanti, Andershed, & DeLisi, 2017). Research shows that the prevalence of psychiatric disorders among the population in juvenile justice institutions (JJIs) is high (e.g. Colins, Vermeiren, Vreugdenhil, van den Brink, Doreleijers, & Broekaert, 2010). Hence, thorough screening and assessment is key and should be both comprehensive and inclusive.

Next to the subjects' characteristics or behavior, variety in diagnostic assessment is influenced by rater characteristics, such as the relationship with the subject, psychopathology of the rater or time spent with the subject (Smith, 2007). Researchers suggest that discrepancies in behavior across settings are however meaningful (Achenbach et al. 1987), and are partly at the root of inter-rater differences in reports of symptomatology. Rather than considering symptoms of psychopathology as generalized traits, which could mean these discrepancies signal a lack of reliability, this variety can yield important and clinically meaningful information, and thus 'should be embraced' (Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). Therefore, instead of considering one sole method or source of information as a golden standard, which is common in for example medical sciences (e.g. blood pressure meter), in psychiatry, the collection of data from multiple sources can be seen as the desired standard (Hunsley & Mash, 2007, Colins et al., 2008). Especially when assessing a culturally diverse group, this multifaceted approach is key.

The screening of mental health needs of detained youths worldwide has immensely improved over the last decade after the introduction of screening instruments such as the Massachusetts Youth Screening Instrument, Version 2 (MAYSI-2, Grisso & Barnum, 2000) and the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1997). In the Netherlands, these inventories are part of a routine clinical assessment implemented by the Dutch Ministry of Safety and Justice (Markus, Colins, Vahl, Matser, & Vermeiren, 2009) administered in the first days after entering the JJI. It is aimed at screening mental health needs and targeting and timing interventions, for example when juveniles report suicidal ideations or highly aggressive tendencies. This screening however currently solely relies on self report, of which we know cultural sensitivity is an issue. Also, dossiers are often still lacking or incomplete in the first days or weeks, leaving instruments more aimed at assessing risk, such as the SAVRY, not yet usable.

Consequently, a more comprehensive and inclusive assessment is of great importance. To complement self report, both researchers and practitioners suggested structured observation by staff as a source of diagnostic information (Colins et al., 2008; Hintze, 2005; McCann, 1997; Platzman et al., 1992; Spaans, Barendregt, Haan, Nijman, & de Beurs, 2011; Volpe, DiPerna, Hintze, & Shapiro, 2005). In line with these recommendations, an Observation Checklist (OC) for use by group workers in two JJIs in the Netherlands was developed and implemented. This process of development and implementation is described in detail elsewhere (Lampe, Mulder, Colins, & Vermeiren, 2017). The OC captures six concepts: Proactive Aggression, Reactive Aggression, Impulsivity, Hyperactivity, Signs of Depressed Mood and Lack of Reciprocity in Contact.

When performing risk assessment in the JJI, such as when using the SAVRY, all sources available are consulted. Both the self reports and the OC can yield important information to contribute to this aim, and have the advantage of taking place in the first days and weeks of the incarceration, providing an unique opportunity to help an early assessment of risk and treatment needs. This is especially important, because in the Netherlands large groups of youths already leave the institution again after the first court hearing, around ten days after incarceration. Risk evaluation regularly takes place in an outpatient setting, but often does use data from the JJI. Moreover, as the population is culturally diverse and we are aware of possible biases in assessment of this group, it is essential to explore the possible influence of ethnicity in these 'building blocks' of a wider (risk) assessment.

This paper aims to contribute to the discussion on cultural sensitivity in the assessment of incarcerated, multicultural youths. To this end, we explored the overlap and differences between observation data collected with the OC and information from self report, in a cultural diverse sample of detained youths. As described above and based upon the knowledge that questionnaires containing similar questions diverge between sources (e.g. parent and child), we expect discrepancies between these measures. In concordance with suggestions mentioned, we anticipate on yielding valuable information on cultural sensitive assessment by contemplating on similarities and differences between these measures, and the role of ethnicity in these differences or similarities. First, we are interested in the general scores of our diverse sample on the measures, e.g. the OC and the self reports, zooming into possible differences between ethnic backgrounds. Second, and highly important for clinical use, we aim to test whether detained (groups of) youths of different ethnic backgrounds could be identified that differ in the discrepancies between self report and observation; e.g. which juveniles score high on self report and high on the same behaviour rated in the OC or the other way around. As we know that juveniles of some ethnic backgrounds

tend to report fewer problems than their peers of Dutch origin, we are specifically interested whether the two measures diverge more in this subgroup. To this end, we use self report data collected over the same period as the observation took place.

Methods

Subjects

Data were collected as part of a standardized mental health screening and assessment in one centrally located JJI in the Netherlands. Participants were male youths, mostly pre-trial and sometimes after conviction, entering this JJI between February 2013 and September 2014. They were placed in the same influx group, where the average stay was 10 days. Structured observation by group workers was part of each shift and of the daily routine. Youths that were observed for less than 5 shifts were excluded from the study. This led to the exclusion of 64 youths, resulting in a sample size of 371 youths.

A standardized mental health screening was administered to almost each youth entering a Juvenile Justice Institution (JJI) in the Netherlands. Between February 1st 2013 and October 1st 2014, of these 371 male adolescents 257 completed this intake procedure. Finally, another 29 juveniles were excluded because they were 18 years or older, thus exceeding the age range for which the MAYSI-2 and SDQ are developed. The final sample thus consisted of 228 juveniles (M age = 16.42, SD 1.13, range 13-17 years), that were observed and also took part in the mental health screening.

In our sample, ethnic minorities were relatively overrepresented compared to the general population in the Netherlands, which is common in Dutch JJI's (Veen, Stevens, Andershed, Raaijmakers, Doreleijers, & Vollebergh, 2011). Of our group of 228 juveniles, the majority, 30.2%, was of Moroccan descent. Another 11.2% was of Dutch origin, 11.2% of Surinamese descent, 9.1% had an Netherlands Antillean background and 8.2% was of Turkish origin. Other ethnic groups of youths, or adolescents of whom data were missing regarding their ethnicity, constituted 30.1% percent of the final sample.

Measures

Massachusetts Youth Screening Instrument-Second Version (MAYSI-2)
The MAYSI-2 (Grisso and Barnum, 2000) is a screening tool that was developed to use with detained youths aged 12-17 years. This self report instrument can be administered by non-clinicians and it contains 52 dichotomous yes/no items on the presence of a wide variety of emotional, behavioral and psychological symptoms or behaviors

experienced in the past few months. Administration takes around 15 minutes. Research suggests the Dutch version of the MAYSI-2 also provides a reliable screening of mental health needs (Colins et al., 2015). In this study, only the depressed-anxious (nine items; α= .66) scale was used.

Strength and Difficulties Questionnaire (SDQ)
The SDQ (Goodman, 1997; van Widenfelt et al. 2003) is a self report tool that screens the psychosocial functioning of children and adolescents aged 11-16. It was designed for use in the general population (Goodman, 2001), but has previously been used in juvenile justice populations (Vahl, Colins, Lodewijks, Markus, Doreleijers, & Vermeiren, 2014). Each scale has three response categories (not true=0, somewhat true=1, certainly true=2). In this study, only the Hyperactivity (α=.66) and Conduct Problems (α= 0.47) subscales of the self report inventory were used, each containing five items.

Reactive Proactive Aggression Questionnaire (RPQ)
The self report version of the RPQ (Raine et al., 2006; Cima, Raine, Meesters, & Popma, 2013) contains 23 items and is used to examine reactive and proactive aggression in both youths and adults. Proactive aggression (α =.85) is assessed by 12 items and the other 11 items assess reactive aggression (α =.86). Answers range from 'never' to 'sometimes' or 'often' and score respectively 0, 1 or 2 points. The internal consistency and validity of RPQ scores in detained male adolescents in the Netherlands are good to excellent (Colins, 2015).

Youth Psychopathic traits Inventory (YPI)
The YPI (Andershed, Kerr, Stattin, & Levander, 2002) is a 50-item self report questionnaire designed to measure psychopathic-like traits in adolescents aged 12 years and up. The Dutch version of the YPI is found to be internally consistent, and correlations with e.g. aggression and conduct problems, support the convergent validity in detained male adolescents (Colins, Fanti, Andershed, Mulder, Salekin, Blokland, & Vermeiren, 2017). The YPI is organized into three dimensions, an interpersonal (α = .89), affective or callous-unemotional (α = .77), and behavioral/lifestyle dimension (α = .86). The wording of the items is designed in such a way that psychopathic-like traits seem positive qualities. Each item in the YPI is scored on a 4-point Likert scale ranging from "Does not apply at all" to "Applies very well". In this study, only 'Impulsivity' is used, which is part of the behavioral/lifestyle dimension, and is based on three items.

The Observation Checklist (OC)
This structured observation checklist for use by group workers was developed for Dutch JJIs and implemented in the influx group in February 2013. Scoring takes place

on a three-point scale. A score of 0 indicates that the concept of interest has not been observed, whereas a score of 1 indicates that the concept only occurred once with a light intensity. A score of 2, finally, indicates that the concept occurred more than once or only once but with clearly negative consequences for the youth, the group or others. The scoring guidelines are explained in the manual and practiced during the two-day training that proceeded the implementation on the influx group where juveniles entered the JJI. Group workers were asked to fill in the OC in consultation with their co-group workers at the end of each shift. Psychometric evaluation demonstrated acceptable to excellent IRR, when expressed as percentage of agreement (Lampe, Mulder, Vermeiren, & Colins, 2023), and the aggression scales have been found of predictive value for incidents later on in the institution (Lampe, Kok, van der Lans, Mulder, & Vermeiren, in press).

Ethnic Background

We use the definition of ethnicity provided by the Dutch Central Bureau of Statistics (2012), that categorizes someone as from a specific ethnic background (e.g. Turkish) when he or she, or (at least) one of the parents is born in another country (e.g. Morocco). When parents differed in country of birth, the mother's country of birth is used to determine the child's ethnicity. Subjects were classified as Dutch when the juvenile and both parents were born in the Netherlands. All others were categorized as of mixed origin.

Procedure

Subjects were observed and subsequently rated using the OC directly after entering the institution by group workers on the influx group. Subjects were rated twice daily on weekdays at the end of every shift (.i.e. an early and a late shift). In weekends, because there is only one shift for the day the subjects were rated once. Subjects filled in the MAYSI-2, SDQ, RPQ and YPI as part of a standardized screening procedure for mental health problems. Generally, screening took place within the first couple of days after entering the JJI. Oral and written information about the aims and content of the screening procedure was handed to all subjects to make them aware that the information would be used to provide the best matched care to their, possible, mental health needs. During the screening, JJI personnel was available to answer potential questions or to assist in the administration. As this screening was meant for clinical use, this routine procedure did not include confidentiality and anonymity guarantees. In concordance with Dutch law, informed consent was not required since aggregated, anonymized data was used, collected from the juveniles' own clinical assessment. Passive informed consent was obtained through standardized information provided by the JJIs upon start of detention; youths and their parents were informed that the

mental health screening and assessment outcomes would be, anonymously, used for scientific research, unless they declined. The Medical Ethical Review Board of the Leiden University Medical Center (LUMC) certified that current study was not subject to the applicable act (the Medical Research Involving Human Subjects Act; In Dutch: Wet Medisch wetenschappelijk Onderzoek met mensen, WMO).

Data-handling

Parallel constructs: matching OC and self report data

OC concepts were matched to relevant self report scales measuring parallel constructs. The OC concept Lack of Reciprocity in Contact did not match any of the available self report data concepts and was not used for comparison. To be able to separately value the observation of aggression, apart from type, we constructed a composite score of aggression using the OC data. We matched this composite score to the RPQ total aggression score. Because the RPQ inquires about aggressive behavior as a trait (and not, for example, behavior in the last few months) we were also interested in a more concrete measure. Therefore the composite aggression score was also matched to the SDQ conduct problem score.

Statistics

Data were collected and then analysed using SPSS 24.0. Of the observation data, means (M) per concept were calculated approximately over a time span of maximum four weeks. The mean amount of observations was 16.4, ranging from 5 to 44 observations with a standard deviation (SD) of 10.2. For the self report scales, sumscores were used.

Descriptive statistics (N, M and sd) were calculated, because of very different scales between the OC and self report, we used Z-scores. For each OC concept and each self report scale, possible differences between ethnic groups were explored. Because the data are not normally distributed, a Kruskal Wallis test was performed, followed by pairwise comparisons using the Dunn-Bonferroni post-hoc multiple analyses correction. In order to determine whether discrepancies between self report and the OC differed between ethnic subgroups, Z-scores were calculated and subtracted from each other, SR - OC. Descriptive analyses were performed. As the normality assumption for ANOVA was again violated, and outliers were visually identified, once more a Kruskal Wallis was performed, followed by a Dunn-Bonferroni correction.

Results

Descriptives

Because scales of self report and OC differ, Z-scores on the self report subscales and OC concepts were calculated, as presented in Figure 1, giving a clear overview of the different scoring patterns among groups with varying ethnic backgrounds. Other descriptives are available on request.

Figure 1 Z-scores of means of OC and selfreport scales per ethnic group

Ethnic group differences on self report and OC scales

On the OC concepts and on the self report scales, differences between ethnic groups were found, shown in Table 1. The Dutch youths self reported higher scores than the Moroccan youths on almost all scales: Proactive aggression, Reactive aggression, Impulsivity, Hyperactivity and Total aggression. This was not found for Depression/ anxiety. On Depression/anxiety, the ethnic Surinamese youths scored higher than the ethnic Moroccan youths. On Hyperactivity, the juveniles with Dutch backgrounds reported higher scores than the ones with Turkish backgrounds as well, and the youths with Surinamese backgrounds reported higher scores than the youths with Moroccan and Turkish roots.

Regarding the scores on the OC, an opposite trend was found. On Reactive aggression, Hyperactivity and the composite score of Total aggression, youths of Moroccan

origin were rated higher on the OC than the group juveniles from 'other' ethnic origin. Juveniles with a Moroccan background were also rated higher on Impulsivity that were youths of Turkish origin. The concept Signs of depressed mood was observed and rated higher in the Dutch youths than in the Moroccan youths.

Table 1 Ethnic group comparison on OC and on self report scales (pairwise compared by Kruskal Wallis, using Dunn-Bonferrroni correction for multiple tests)

Observation checklist data	Adj P<0.1 (2-sided, Bonferroni corrected)	H(df=5)	Adj. p	Self report data	Adj P<0.1 (2-sided, Bonferroni corrected)	H(df=5)	Adj. P
Proactive aggression	none			**Proactive Aggression (RPQ)**	D>M	8.204	0.063
Reactive aggression	M>O	10.937	0.014	**Reactive Aggression (RPQ)**	None		
Impulsivity	M>T	7.850	0.076	**Impulsivity (YPI)**	D>M	8.662	0.049
Hyperactivity	M>O	8.818	0.045	**Hyperactivity (SDQ)**	D>M D>T S>T S>M	10.753 8.777 7.351 9.031	0.016 0.046 0.049 0.023
Signs of depressed mood	D>M	9.022	0.400	**Depression-anxiety (MAYSI)**	S>M O>M	17.897 7.381	0.000 0.099
Lack of reciprocity in contact	none						
Aggression composite score	M>O	36.408	0.007	**Conduct problems (SDQ)** **Total Aggression (RPQ)**	None D>M		

Note. Adj = adjusted, P = level of significance, df=degrees of freedom, H= Kruskal Wallis test statistic, M= Moroccan origin, D= Dutch origin, S= Surinamese origin, T=Turkish origin, O= other unspecified origin.

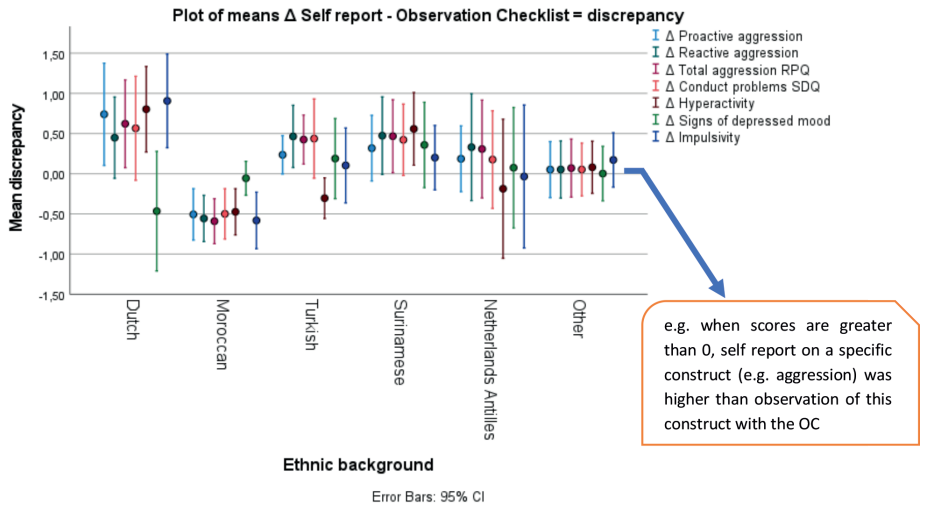
Ethnic group differences in the discrepancy between OC and self report scales

Exploration on the subtraction SR - OC, revealed different significant findings. In Figure 2 are the means depicted of the subtraction scores when subtracting the Z-scores OC from SR (e.g. ZSR - ZOC), organized by ethnicity. Our Kruskal Wallis test of independent samples, that uses the median, rejected the hypotheses that the distribution of discrepancies between self report and concordant OC scales are the same across different ethnic backgrounds. Differences in discrepancies self report-OC among ethnic origins were found when they were pairwise compared, using the adjusted significance level corrected by Bonferroni at p<.05.

Although Figure 2 is based on means and Kruskal Wallis uses the median, it gives a good overview of the differences between ethnic groups, that we further analysed by the Kruskal Wallis. On all matched constructs (e.g. OC concept to parallel self report scale) except Signs of depression and Depression / anxiety, youths from Moroccan background report less problems (on the self report) than are observed, whereas for the ethnic Dutch, this is the reverse: they report more problems themselves, than are observed and rated on the OC. For Proactive aggression (H(5)= 56.733, p=0.003), Reactive aggression (H(5)=51.326, p=0.012), Impulsivity (H(5)=68.381, p=0.000), Hyperactivity (H(5)= 62.490, p=0.001), Total aggression (SDQ conduct problems, H(5)=46.305, p=0.038) and Total aggression (RPQ total H(5)=59.448, P=0.002), youths with Moroccan background differed from ethnic Dutch. On all the same scales except Impulsivity, the youths with Moroccan background also differed from the youths with a Surinamese background with the following values for Proactive aggression H(5)=-52.720, p=0.007, Reactive aggression H(5)=-61.288, p=0.010, Hyperactivity H(5)=-59.794, p=0.001, Total aggression (SDQ conduct problems) H(5)=-45.745, p=0.043 and Total aggression (RPQ total) H(5)=-62.997, p=0.001. The mixed, other group also differed from the youths with a Moroccan background on Reactive aggression (H(5)=33.711, p=0.037), Impulsivity (H(5)=-35.900, p=0.020) and Total aggression (RPQ total H(5)=36.168, p=0.017). Moroccan-ethnic youth differed from youths with Turkish roots on Reactive aggression (H(5)=-57.607, p=0.014) and on Total aggression (RPQ total, H(5)=-60.768, p=0.007). Juveniles with a Turkish background differed from ethnic Dutch youths on Hyperactivity, H(5)=60.206, p=0.046.

The Signs of depression concept is the exception, juveniles with Dutch backgrounds are more often observed as showing signs of depressed mood, than they report themselves, but no differences between ethnicities in the discrepancies were found.

Figure 2 Means of the subtraction of Z scores of self report minus OC



Discussion

Our data revealed relevant findings on the role of cultural factors in observation and self report, both important sources for risk assessment and tailored treatment to prevent recidivism. We found that ethnic background plays a significant role in self report and observation, and in how these measures diverge and converge, underscoring the importance for a multimethod approach in (risk) assessment and a cultural sensitivity when interpreting information.

First, between ethnic-group differences were found in the self report data. Youths of Moroccan background self reported fewer problems than youths with a Dutch background on the more externalizing scales: Aggression, Impulsivity and Hyperactivity. On Depression/anxiety they also showed the lowest mean scores, but only significantly different (lower) than the youths with a Surinamese background. On Hyperactivity, the juveniles with Dutch background reported higher scores than those with Turkish background as well. Youths with a Surinamese background reported higher scores than youths with Moroccan and Turkish roots.

Second, on the OC, differences were found between scores of ethnic subgroups, but in opposite directions than in self report. Again, the scores of youths of Moroccan origin differed from the other ethnic backgrounds, more than other ethnic groups did. Here a clear trend was visible: externalizing concepts, e.g. Total aggression and

Reactive aggression, were observed more in Moroccan ethnic youth than in the group of 'other' origin. In contrast, internalizing behavior, e.g. Signs of depressed mood was observed and rated higher in the youths from a Dutch background than in the youths of Moroccan origin.

When reviewing the two sources together, it becomes very clear how sources of information can diverge depending on ethnicity, but also on the construct measured. Our most prominent finding is how youths from Moroccan and Dutch background diverge in the relation between self report and observation. Youths from Moroccan origin report less problems than are observed, whereas ethnic Dutch youths report more problems than are observed. Interestingly, for Signs of depressed mood combined with Depression / anxiety, this trend does not exist; observers rate more depressive signs than ethnic Dutch juveniles report themselves- but no differences between ethnicities in the discrepancies were found.

The first two results, those that show how youths from different ethnic backgrounds vary in their ratings on self report and on observation, add up to our third result: ethnicity influences how both sources of information diverge and converge.

When we look at the reliability of self report in different ethnicities, previous studies give some direction in how we should interpret our results. For example, Batenburg van Batenburg-Eddes, Butte, van de Looij-Jansen, Schiethart, Raat, de Waart, & Jansen (2012) found that in youth of Moroccan origin a greater divergence existed between self report and police data, with youths of Moroccan origin less likely to self report police contact conform the data. The lower scoring pattern of youths with Moroccan background on self report is in our study also the most striking, and is in line with previous research that found lower scores tendencies in self report to be more common in immigrants (Davies and McKelvey (1998) and youths with Moroccan background in the Netherlands (Veen, Stevens, Andershed, Raaijmakers, Doreleijers, & Vollebergh, 2011; Colins, 2016). Possible explanations are cultural differences in how psychopathology is defined, different social cultural expectations, shame or fear of judicial concequences (Davies and McKelvey, 1998;, van Batenburg-Eddes et al. 2012; Crone, Bekkema, Wiefferink & Reijneveld, 2009; Sue, 1995). Considering the latter, in the Netherlands, people of Moroccan descent are relatively often victims of discrimination and of racial profiling by police (Amnesty International, 2013), feeding distrust of authorities and institutions, perhaps contributing to not being open on self reports. Research also pointed to possible bias in the Dutch court system, with ethnic Dutch juveniles held more often diminished responsible although mental disorders were found similar frequent, and, after pre trial evaluation, less often advised to be placed in

a JJI (Vinkers & Duits, 2011). The latter explanation is also given by Veen and colleagues (2011), who mention disparities in sentencing in the Dutch court system, leading to the incarceration of relatively less troubled youth of Moroccan descent, compared to their Dutch-background peers. They imply that the lower scores on self report are reliable and means they have less mental problems. However, other researchers do not find less mental health problems in this group (Vinkers and Duits, 2011). Remarkable is how youths with Surinamese and to a lesser extent those with a Netherlands Antilles background, have similar scoring tendencies as the youths with an all Dutch origin. The shared history of these cultures, as both were colonized by the Netherlands in the 17th century and thus dates back centuries, is perhaps an explanation for more similarities in aforemost language, but also awareness, exposure and expressing of complaints or feelings. The youths with Turkish backgrounds have more similarities in history and religion with the juveniles with Moroccan background, often children of (grand)parents that were recruited to work in the Netherlands in the 1960s, at first often with the intention to return (Van Meeteren, Van de Pol, Dekker, Engbersen & Snel, 2013). Earlier research has found the similar clusters in answering tendencies when researching self reported crime, e.g. youth with Turkish and Moroccan backgrounds self reporting less police contacts, whilst juveniles with Dutch, Netherlands Antillean and Surinamese background reported more police contacts (van Batenburg-Eddes et al, 2012; Junger, 1989). Scoring profiles of youths of Turkish and Moroccan background are similar on the self report of Hyperactivity, but do differ on the other aspects.

On the observation checklist, group workers are rating the juveniles on the group. There is a possibility that in fact, youths of Moroccan descent show more aggression and thus it is rated more often. However, alternative plausible explanations also exist. First, not registered but noticed, most group workers were of Dutch origin. Research shows that ethnic background of the subject influences perception of emotional expressions, e.g. people from a certain ethnic background recognize faces and expressions more adequate in people from the same origin, than in cross ethnic groups (Lipp, Craig & Dat, 2015; Bijlstra, Holland, & Wigboldus, 2010). Perhaps the elevated scores of Signs of depressed mood, as are observed in the ethnic Dutch group but barely in the other groups, can be understood in this direction. Second, racial stereotypes and prejudices have to be considered when interpreting our results. It is remarkable how externalizing behavior is observed more in the Moroccan group, both in contrast with other groups and in contrast with internalizing problems, and future research should aim at detangling this further, for example by combining observation and self report data with institutional incidents or recidivism data. Previous research found in line with our findings, found that teachers rated internalizing behavior less often in Moroccan than Dutch background boys, and externalizing behavior more often (Vollebergh, ten

Have, Dekovic, Oosterwegel, Pels, Veenstra, ... & Verhulst, 2005). Third, it is commonly accepted that how internalizing disorders are expressed or presented differs among cultures (Kirmayer, 2001). It should be taken into account the explanation that youth with an immigrant background, can use somatic complaints as an expression of mental troubles (Bengi-Arslan, Verhulst, & Crijnen, 2002), or use more externalizing gestures to express distress or (agitated) depression. Incorporating frequency of use of medical services into future research, could perhaps shed more light on somatic complaints as a sign of distress.

Our results contribute to this complex discussion by combining observation and self report data, and showing that differences exist between ethnic backgrounds and internalizing and externalizing scales. these converge more regarding internalizing problems, but diverge more concerning the more externalizing scales. It becomes clear that youth with Moroccan backgrounds differ largely from the other groups in both sources of information. In the case of risk assessment or tailoring treatment, we believe our findings underscore the urgency not to take into account only one source of information. Taking the aforementioned literature together, we believe in all mentioned explanations some truth is hidden and they all add to the equation. Most likely a complex dynamic exists between underreporting on self reports, especially of 'unfavourable' behavior, a different way and language of expressing emotions, a lack of cross-cultural validation of self report, biases in the Dutch judiciary and police system leading to earlier incarceration of (less mentally disturbed) immigrant groups, and biases in observation when rating behavior in the juvenile justice institution. Despite all these limitations, we are convinced a multi faceted approach in (risk) assessment is key. In concordance with others (Hunsley & Mash, 2007, Colins et al., 2008), we believe that the collection of data from multiple sources should be best practice, as our findings illustrate. Contemplation on the meaning of discrepancies- both in practice and in research- and incorporating cultural sensitivity in this matter is of urgent importance. Knowledge about the role ethnic background plays, in subject and observer, and a multi-method assessment yields a more thorough and tailored evaluation.

Some limitations should be addressed. Groups of ethnic background are relatively small, making it harder to draw conclusions on these groups. However, crucial significant findings were found, underscribing the power and significance of the reported differences. The sparseness of differences with the biggest group, e.g. 'other or unknown ethnicity', seems logical as it is presumably a very mixed group, evening out differences we found in the other groups. This brings up another limitation: the definition of ethnicity we used, assigned third generation immigrants to the native

Dutch group, even though it is likely they self-identify their cultural identity as that of their grandparents, and also speak their grand parents' language at home (Stronks, Kulu-Glasgow & Agyemang, 2009). Having youths self-identify their cultural identity, could have perhaps refined the results further. Our final limitation concerns the limited researched reliability of the Observation Checklist. The finding that aggression scales of the OC are of predictive value for incidents later on in the institution, is however promising (Lampe, Kok, van der Lans, Mulder and Vermeiren, in press).

Next to already mentioned directions of research, we believe that a further focus on the role cultural factors play in (risk) assessment is necessary, and should be also extended to other sources of information, such as parents and teachers- next to observation and self report-, in this dynamic and complex group. Furthermore, it is of importance to stress that current findings only concern males- as crime numbers increase in females but cultural dynamics can also differ between genders (Leiber and Peck, 2015), more research into the role of gender and culture is in place.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological bulletin, 101(2)*, 213.

Adriaanse, M., van Domburgh, M. D. L., Veling, W., & Doreleijers, T. A. H. (2011). Psychische problemen en stoornissen bij Marokkaans Nederlandse kinderen en jongeren. Eindrapportage van een onderzoek naar de aard en omvang van emotionele, gedrags-en andere psychische problemen en stoornissen en risico-en beschermende factoren bij Marokkaans Nederlandse kinderen en jongeren.

Amnesty International. (2013). *Proactief politieoptreden vormt risico voor mensenrechten* [Proactive police performance forms risk for human rights]. Retrieved from https://www.amnesty.nl/actueel/proactief-politieoptreden-leidt-tot-discriminatie-in-nederland

Andershed, H., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: initial test of a new assessment tool. In: E. Blaauw & L. Sheridan (Eds.), *Psychopaths: current international perspectives* (pp 131-158). Elsevier, The Hague.

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.

Bengi-Arslan, L., Verhulst, F. C., & Crijnen, A. A. (2002). Prevalence and determinants of minor psychiatric disorder in Turkish immigrants living in the Netherlands. *Social psychiatry and psychiatric epidemiology, 37*(3), 118-124.

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology, 46*(4), 657-663.

Bishop, D. M. & Frazier, C. E. (1996). Race effects in juvenile justice decision-making: Findings of statewide analysis. *Journal of Criminal Law and Criminology 86(2)*, 392-414.

Colins, O. F. (2016). Assessing reactive and proactive aggression in detained adolescents outside of a research context. *Child Psychiatry & Human Development, 47*(1), 159-172.

Cima, M., Raine, A., Meesters, C., & Popma, A. (2013). Validation of the Dutch Reactive Proactive Questionnaire (RPQ): differential correlates of reactive and proactive aggression from childhood to adulthood. *Aggressive Behavior, 39(2)*, 99-113.

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. *Journal of Clinical Psychiatry, 69(1)*, 141.

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. *Canadian Journal of Psychiatry, 55*(4), 255-263.

Colins, O., Vermeiren, R. R. J. M., 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.

Colins, O. F., Grisso, T., Vahl, P., Guy, L., Mulder, E., Hornby, N., ... & Vermeiren, R. (2015). Standardized screening for mental health needs of detained youths from various ethnic origins: the Dutch Massachusetts Youth Screening Instrument-Second Version (MAYSI-2). *Journal of psychopathology and behavioral assessment, 37(3)*, 481-492.

Colins, O. F., Damme, L. V., Andershed, H., Fanti, K. A., & DeLisi, M. (2017). Self reported psychopathic traits and antisocial outcomes in detained girls: A prospective study. *Youth Violence and Juvenile Justice, 15(2)*, 138-153.

Crone, M., Bekkema, N., Wiefferink, C., & Reijneveld, S.A. (2009). Professional identification of psychosocial problems among children from ethnic minority groups: room for improvement. *The Journal of Pediatrics*, (156). 277-284.

Davies, L. C., & McKelvey, R. S. (1998). Emotional and behavioural problems and competencies among immigrant and non-immigrant adolescents. *Australian and New Zealand Journal of Psychiatry*, *32*(5), 658-665.

Desai, R. A., Falzer, P. R., Chapman, J., & Borum, R. (2012). Mental illness, violence risk, and race in juvenile detention: Implications for disproportionate minority contact. *American Journal of Orthopsychiatry, 82(1)*, 32-40.

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual Research Review: Embracing not erasing contextual variability in children's behavior-theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 53(5), 558-574.

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

Fite, P. J., Raine, A., Stouthamer-Loeber, M., Loeber, R., & Pardini, D. A. (2010). Reactive and proactive aggression in adolescent males: Examining differential outcomes 10 years later in early adulthood. *Criminal Justice and Behavior, 37(2)*, 141-157.

Florsheim, P., Shotorbani, S., Guest-Warnick, G., Barratt, T., & Hwang, W. C. (2000). Role of the working alliance in the treatment of delinquent boys in community-based programs. *Journal of clinical child psychology, 29(1)*, 94-107.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry, 38(5)*, 581-586.

Grisso, T., & Barnum, R. (2000). Massachusetts Youth screening instrument, second version (MAYSI-2): User's manual and technical report. *Worcetser, MA: University of Massachusetts Medical School.*

De Haan, A. M., Boon, A. E., Vermeiren, R. R., & De Jong, J. T. (2012). Ethnic differences in utilization of youth mental health care. *Ethnicity & health, 17(1-2)*, 105-110.

Hintze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review*, 34, 507-519.

Hillege, S., Das, J., & de Ruiter, C. (2010). The Youth Psychopathic Traits Inventory: Psychometric properties and its relation to substance use and interpersonal style in a Dutch sample of non-referred adolescents. *Journal of Adolescence*, *33*(1), 83-91.

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

Janssen, M. M. M., Verhulst, F. C., Bengi-Arslan, L., Erol, N., Salter, C. J., & Crijnen, A. A. M. (2004). Comparison of self reported emotional and behavioral problems in Turkish immigrant, Dutch and Turkish adolescents. *Social Psychiatry and Psychiatric Epidemiology: The International Journal for Research in Social and Genetic Epidemiology and Mental Health Services*, *39*(2), 133-140.

Junger, M. (1989). Discrepancies between police and self report data for Dutch racial minorities. *The British Journal of Criminology*, *29*(3), 273-284.

Kirmayer, L. J. (2001). Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of clinical psychiatry, 62*, 22-30.

Ladd, G. W., & Kochenderfer-Ladd, B. (2002). Identifying victims of peer aggression from early to middle childhood: analysis of cross-informant data for concordance, estimation of relational adjustment, prevalence of victimization, and characteristics of identified victims. *Psychological assessment, 14(1)*, 74.

Lampe, K. G., Mulder, E. A., Colins, O. F., & Vermeiren, R. R. (2017). The inter-rater reliability of observing aggression: A systematic literature review. *Aggression and violent behavior*, 37, 12-25.

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.

Lampe, K., Kok, M., van der Lans, R., Mulder, E., & Vermeiren, R. (in press). Risk assessment in incarcerated adolescents-Using an observation checklist to predict aggressive incidents. *Journal of Forensic Psychology Research and Practice.*

Leiber, M. J., & Peck, J. H. (2015). Race, gender, crime severity, and decision making in the juvenile justice system. *Crime & Delinquency, 61*(6), 771-797.

Lipp, O. V., Craig, B. M., & Dat, M. C. (2015). A happy face advantage with male caucasian faces: It depends on the company you keep. *Social Psychological and Personality Science*, *6*(1), 109-115.

Markus, M., Colins, O., Vahl, P., Matser, D., & Vermeiren, R. (2009). Massachusetts Youth Screening Instrument-Second Version- Gauthoriseerde Nederlandse vertaling.

McCann, J. J. (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*, 63-72.

Van Meeteren, M., Van de Pol, S., Dekker, R., Engbersen, G., & Snel, E. (2013). Destination Netherlands. History of immigration and immigration policy in the Netherlands. *Immigrants: Acculturation, socio-economic challenges and cultural psychology*, 113-170.

Muir, N. M., Viljoen, J. L., Jonnson, M. R., Cochrane, D. M., & Rogers, B. J. (2020). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) with Indigenous and Caucasian female and male adolescents on probation. *Psychological Assessment*, 32(6), 594-607.

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.

Polman, H., de Castro, B. O., Koops, W., van Boxtel, H. W., & Merk, W. W. (2007). A meta-analysis of the distinction between reactive and proactive aggression in children and adolescents. *Journal of Abnormal Child Psychology, 35*, 522-535.

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., ... & Liu, J. (2006). The reactive-proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive behavior*, 32(2), 159-171.

Shepherd, S. M., Luebbers, S., Ferguson, M., Ogloff, J. R., & Dolan, M. (2014). The utility of the SAVRY across ethnicity in Australian young offenders. *Psychology, Public Policy, and Law*, 20(1), 31.

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.

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*, 374-378.

Stronks, K., Kulu-Glasgow, I., & Agyemang, C. (2009). The utility of 'country of birth'for the classification of ethnic groups in health research: the Dutch experience. *Ethnicity & health*, *14*(3), 255-269.

Vahl, P., Colins, O. F., Lodewijks, H. P., Markus, M. T., Doreleijers, T. A., & Vermeiren, R. R. (2014). Psychopathic-like traits in detained adolescents: Clinical usefulness of self report. *European child & adolescent psychiatry*, *23*(8), 691-699.

Veen, V., Stevens, G., Doreleijers, T., Van Der Ende, J., & Vollebergh, W. (2010). Ethnic differences in mental health among incarcerated youths: do Moroccan immigrant boys show less psychopathology than native Dutch boys?. *European child & adolescent psychiatry*, *19*(5), 431-440.

Veen, V. C., Stevens, G. W., Andershed, H., Raaijmakers, Q. A., Doreleijers, T. A., & Vollebergh, W. A. (2011). Cross-ethnic generalizability of the three-factor model of psychopathy: The Youth Psychopathic Traits Inventory in an incarcerated sample of native Dutch and Moroccan immigrant boys. *International journal of law and psychiatry*, *34*(2), 127-130.

Vinkers, D. J., & Duits, N. (2011). Crime and mental disorders among native Dutch and ethnic minority juvenile defendants in the Netherlands. *International journal of law and psychiatry*, *34*(2), 131-134.

Vollebergh, W. A., ten Have, M., Dekovic, M., Oosterwegel, A., Pels, T., Veenstra, R., de Winter, A., Ormel, H., & Verhulst, F. (2005). *Mental health in immigrant children in the Netherlands. Social psychiatry and psychiatric epidemiology, 40(6)*, 489-496.

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*, 454-474.

van Widenfelt B.M., Goedhart, A.W., Treffers, P.D.A., Goodman, R. (2003). The Dutch version of the strengths and difficulties questionnaire (SDQ). *European Journal of Child and Adolescent Psychiatry 12(6)*, 281-289.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. 2000!. Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 68*, 1038-1050