Journal of Threat Assessment and Management

Ethnic Differences in Assessment: How Self-Report and Observation Converge and Diverge Among Ethnically Diverse Incarcerated Youths

Kore G. Lampe, Eva A. Mulder, and Robert R. J. M. Vermeiren

Online First Publication, February 2, 2023. https://dx.doi.org/10.1037/tam0000199

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Threat Assessment and Management

ISSN: 2169-4842

https://doi.org/10.1037/tam0000199

Ethnic Differences in Assessment: How Self-Report and Observation Converge and Diverge Among Ethnically Diverse Incarcerated Youths

Kore G. Lampe, Eva A. Mulder, and Robert R. J. M. Vermeiren

Department of Child and Adolescent Psychiatry, Curium-LUMC

While structured observation and self-report (SR) are of great value for risk assessment in forensic youth psychiatry, a number of conclusory variances should be taken into consideration as it pertains to cultural sensitivities. This study therefore researches data collected by self-report using standardized questionnaires, and by an observation checklist (OC) in an ethnically diverse population of incarcerated youths. Our sample consisted of 228 male incarcerated juveniles, with the identifiable majority (30.2%) of Moroccan origin, 11.2% of Dutch origin, 11.2% of Surinamese origin, 9.1% with a Netherlands Antilles origin, and 8.2% of Turkish origin. Adolescents from other origins or whose ethnicity data was missing, constituted 30.1% of the final sample. First, scores on each self-report subscale and OC concept were analyzed for differences between ethnic origins. Second, OC concepts were matched to relevant self-report scales, for example, proactive aggression of the OC to proactive aggression of the Reactive Proactive Aggression Questionnaire. Finally, convergence and divergence between the two matched concepts was analyzed. Large differences were found for the separate methods, and the divergence and convergence between the two methods. Most prominently was the very different scoring profile between youths of Dutch and Moroccan origin, with the latter self-reporting fewer problems than youths of Dutch origin, while more problems were observed. Possible explanations, such as (racial or ethnic) bias in observing, lack of cross-cultural validation for self-report, or biases such as shame, fear of judicial consequences are discussed.

Public Significance Statement

Ethnic differences impact self-report and cultural biases impact observation. This strongly urges practitioners and researchers to use a multimethod approach in risk assessment, which is more culturally sensitive, while being highly sensitive during observation to avoid biases impacting results and outcomes.

Keywords: cultural sensitive assessment, observation, self-report, ethnicity, incarcerated youth

Kore G. Lampe https://orcid.org/0000-0001-6976-2968

The authors wish to thank ZonMw for their generous funding with project No. 15901.0002, and Floor de Heus for her help in the lay out. Funding for this study was provided by ZonMW Grant 15901.0002. ZonMW had no role in the study design, selection, analysis or interpretation of the articles, writing the article, or the decision to submit the article for publication. The authors have no conflicts of interest to disclose. This declaration is read and endorsed by all authors: Kore G. Lampe, Eva A. Mulder and Robert R. J. M. Vermeiren.

The authors confirm that the article has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. The order of authors listed in the article has been approved by all authors.

Correspondence concerning this article should be addressed to Kore G. Lampe, Department of Child and Adolescent Psychiatry, Curium-LUMC, Postbus 15, 2300 AA Leiden, The Netherlands. Email: k.g.lampe@lumc.nl

The 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 observed behavior, features of the setting, and the type of interviewer and instrument (Achenbach et al., 1987; Fazel et al., 2008; Ferdinand et al., 2004; Florsheim et al., 2000). Furthermore, often different methods and sources of information (e.g., parents, youths, teachers, observations by group workers) do not correlate (Colins et al., 2015; Janssen et al., 2004; Smith, 2007; Youngstrom et al., 2000).

Another important factor to consider is the role cultural factors play in assessment. Ethnic minorities are overrepresented in the judiciary system (Bishop & Frazier, 1996; Veen et al., 2011), while assessment methods do not always consider this diverse population. For example, self-report (SR) 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 (Crone et al., 2010; van Batenburg-Eddes et al., 2012). Research also shows that detained youths are 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 structured professional judgment tools such as the Structured Assessment of Violent Risk in Youth (SAVRY) have been found to be of less predictive value for different ethnic groups (Muir et al., 2020; Shepherd et al., 2014).

While ethnic minorities are overrepresented in the judiciary system, they are underrepresented in the Dutch mental health care system (de Haan et al., 2012). Considering the prevalence of psychiatric disorders among ethnic minorities is similar to youths from Dutch origin, it is suggested that psychiatric problems are most probably addressed too late if at all, and only after these juveniles end up in the criminal justice system. Earlier interventions would presumably allow youth from ethnic minorities to access regular mental health services (Adriaanse et al., 2011). This underscores the importance of timely, and above all, culturally sensitive assessment.

Psychiatric disorders elevate the risk for various detrimental outcomes, including recidivism (e.g., Colins et al., 2017; Fite et al., 2009). Research shows that the prevalence of psychiatric disorders among the population in juvenile justice institutions (JJIs) is high (e.g., Colins et al., 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 or observer characteristics, such as the relationship with the subject, psychopathology of the rater or time spent with the subject (Smith, 2007). However, discrepancies in behavior across settings are also suggested to be meaningful (Achenbach et al., 1987) and are partly at the root of interrater 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 et al., 2012). Therefore, instead of considering one method or source of information as a golden standard, which is common for example, in medical sciences (e.g., measuring blood pressure), in psychiatry, the collection of data from multiple sources is seen as the desired standard (). 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 et al., 2009) and are administered in the first days upon entering the JJI. These tools are used to screen for mental health needs and targeting and timing interventions, such as when juveniles report suicidal ideations or highly aggressive tendencies. However, this screening currently relies solely on self-report, which we know is not culturally sensitive. Also, youth's files are often incomplete in the first days or weeks,

which does not allow risk assessing tools such as the SAVRY to be useful at that time. Consequently, a more comprehensive and inclusive assessment of incarcerated youth's mental health needs 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., 2015; Hintze, 2005; McCann et al., 1997; Platzman et al., 1992; Spaans et al., 2011; Volpe et al., 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 et al., in press). 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 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 a unique opportunity for an early assessment of risk and treatment needs. This is especially important, because in the Netherlands large groups of youths leave the institution again after their first court hearing, around 10 days after incarceration. Risk evaluation regularly takes place in an outpatient setting, but often also uses 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 article aims to contribute to the discussion on cultural sensitivity in the assessment of incarcerated, multicultural youth. To this end, we explored the overlap and differences between observation data collected with the OC and the self-reports, in a culturally diverse sample of detained youths. As described above and based upon the knowledge that questionnaires containing similar questions diverge regarding the sources (e.g., parent and child), we expect discrepancies between these measures are expected. In concordance with these, we anticipate on yielding valuable information on cultural sensitive assessment by contemplating on similarities and differences between these measures, and the role of ethnicity potentially plays. First, we are interested in the general scores of our diverse sample on the measures, for example, the OC and the self-reports, zooming into possible differences between ethnic origins. Second, and highly important for clinical use, we aim to test whether detained (groups of) youths of different ethnic origins differ in the discrepancies between self-report and observation; for example, which juveniles score high on self-report and high on the same behavior rated in the OC or the other way around. As we know that youth of certain ethnic origins tend to report fewer problems than their peers of Dutch origin, we are specifically interested in whether the two measures diverge more in this subgroup. To this end, we used self-report data collected over the same period as when the observations took place.

Method

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 pretrial and sometimes after conviction, entering this JJI between February 2013 and September 2014. These youth were placed in the same 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 five shifts were excluded from the study, because when youths enter the institution, they are slowly integrated in the group, leading to less opportunities and less time to observe. When juveniles stay so little time that they are only observed a few shifts, it is safe to assume they have spent a large part of these observations in their cells. We therefore decided not to take such short stays into account. 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 every youth entering a JJI in the Netherlands. Between February 1, 2013 and October 1, 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 is

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, youth from ethnic minorities were relatively overrepresented compared to the general population in the Netherlands, which is common in Dutch JJI's (Veen et al., 2011). Of our group of 228 juveniles, the majority (30.2%) was of Moroccan origin. Another 11.2% was of Dutch origin, 11.2% of Surinamese origin, 9.1% was of Netherlands Antillean origin, and 8.2% was of Turkish origin. Youth with other ethnic origins or for whom this data were missing constituted 30.1% percent of the final sample.

Measures

The 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 3-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 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 they were practiced during the 2-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, and the aggression scales were found of predictive value for later incidents in the institution (Lampe et al., in press).

MAYSI-2

The MAYSI-2 (Grisso & 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 in 15 min by nonclinicians and it contains 52 dichotomous yes/no items regarding the presence of a wide variety of emotional, behavioral and psychological symptoms experienced in the past few months. Research suggests the Dutch version of the MAYSI-2 provides a reliable screening of mental health needs (). Because OC concepts were matched to relevant self-report scales measuring parallel constructs, certain subscales were used. From the MAYSI-2, only the depressed-anxious (nine items; $\alpha = .66$) scale was used

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 et al., 2014) and can also be used in older adolescents such as in our sample (Van Roy et al., 2008). Each scale has three response categories (*not true* = 0, *somewhat true* = 1, *certainly true* = 2). In this study, only the Hyperactivity ($\alpha = .66$) and Conduct Problems ($\alpha = 0.47$) subscales of the self-report inventory were used, each containing five items.

Reactive Proactive Aggression Questionnaire

The self-report version of the Reactive Proactive Aggression Questionnaire (RPQ; Cima et al., 2013; Raine et al., 2006) contains 23 items and is used to examine reactive and proactive aggression in both youths and adults. Proactive aggression ($\alpha = .85$) is assessed by 12 items and the other 11 items assess reactive aggression ($\alpha = .86$). 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, 2016).

Youth Psychopathic Traits Inventory

The Youth Psychopathic Traits Inventory (YPI; Andershed et al., 2002) is a 50-item self-report questionnaire designed to measure psychopathic-like traits in adolescents aged 12 years and older. The Dutch version of the YPI is found to be internally consistent, and correlations with for example, aggression and conduct problems, support the convergent validity in detained male adolescents (Colins et al., 2017). The YPI is organized into three dimensions, an interpersonal

($\alpha = .89$), affective or callous–unemotional ($\alpha = .77$), and behavioral/lifestyle dimension ($\alpha = .86$). The wording of the items is designed in such a way that psychopathic-like traits seem like 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" was used, which is part of the behavioral/lifestyle dimension, and based on three items.

Ethnic Origin

We use the definition of ethnicity provided by the Dutch Central Bureau of Statistics, that categorizes someone as from a specific ethnic origin (e.g., of Turkish origin) when they, or (at least) one of the parents are born in another country (e.g., Turkey). 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

Directly after entering the institution, subjects were observed and subsequently rated using the OC 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). On weekends, because there is only one shift per day, the subjects were only rated once. Subjects filled in the MAYSI-2, SDQ, RPQ and YPI as part of a standardized screening procedure for mental health problems. Generally, this 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 provided to all subjects to inform them that the findings would be used to provide the best matched care to their 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 accordance with Dutch law, informed consent was not required since data were aggregated, anonymized data were used, and it was collected from the juveniles' own clinical assessment. Passive informed consent was obtained through standardized information provided by the JJIs upon the 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 certified that the present study was not subject to the applicable law (the Medical Research Involving Human Subjects Act; In Dutch: Wet Medisch wetenschappelijk Onderzoek met mensen).

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 analyzed using SPSS 24.0. Of the observation data, means (*M*) per concept were calculated approximately over a time span of maximum 4 weeks, taking into consideration how often youths were observed. The mean number 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, and because of very different scales between the OC and self-report, we used Z-scores. For each OC concept and self-report scale, possible differences between ethnic groups were explored. As the data were not normally distributed, a Kruskal–Wallis test was performed, followed by a pairwise

comparison 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 also performed. Due to the normality assumption for analysis of variance was violated again, and outliers were visually identified, a Kruskal–Wallis was performed once more, followed by a Dunn–Bonferroni correction.

Results

Descriptives

Because scales of self-report and OC differed, 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 ethnic groups. Other descriptives are available upon request.

Ethnic Group Differences on Self-Report and OC Scales

Differences between ethnic groups were found regarding the OC concepts and the self-report scales, shown in Table 1. The Dutch origin reported higher scores than the youths of Moroccan origin on almost all scales: Proactive aggression, Reactive aggression, Impulsivity, Hyperactivity, and Total aggression. This was not found for Depression/anxiety. On Depression/anxiety, the youths of Surinamese origin scored higher than the youths of Moroccan origin. On Hyperactivity, the juveniles of Dutch origin reported higher scores than those with Turkish origins, and the youths with Surinamese origins reported higher scores than the youths of Moroccan and Turkish origin.

Regarding the scores on the OC, an opposite trend was found. On reactive aggression, hyperactivity and the composite score of aggression, youths of Moroccan origin were rated higher on the OC than the group juveniles from "other" ethnic origins. Juveniles of Moroccan origin 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 youths of Dutch origin than in the youths of Moroccan origin.

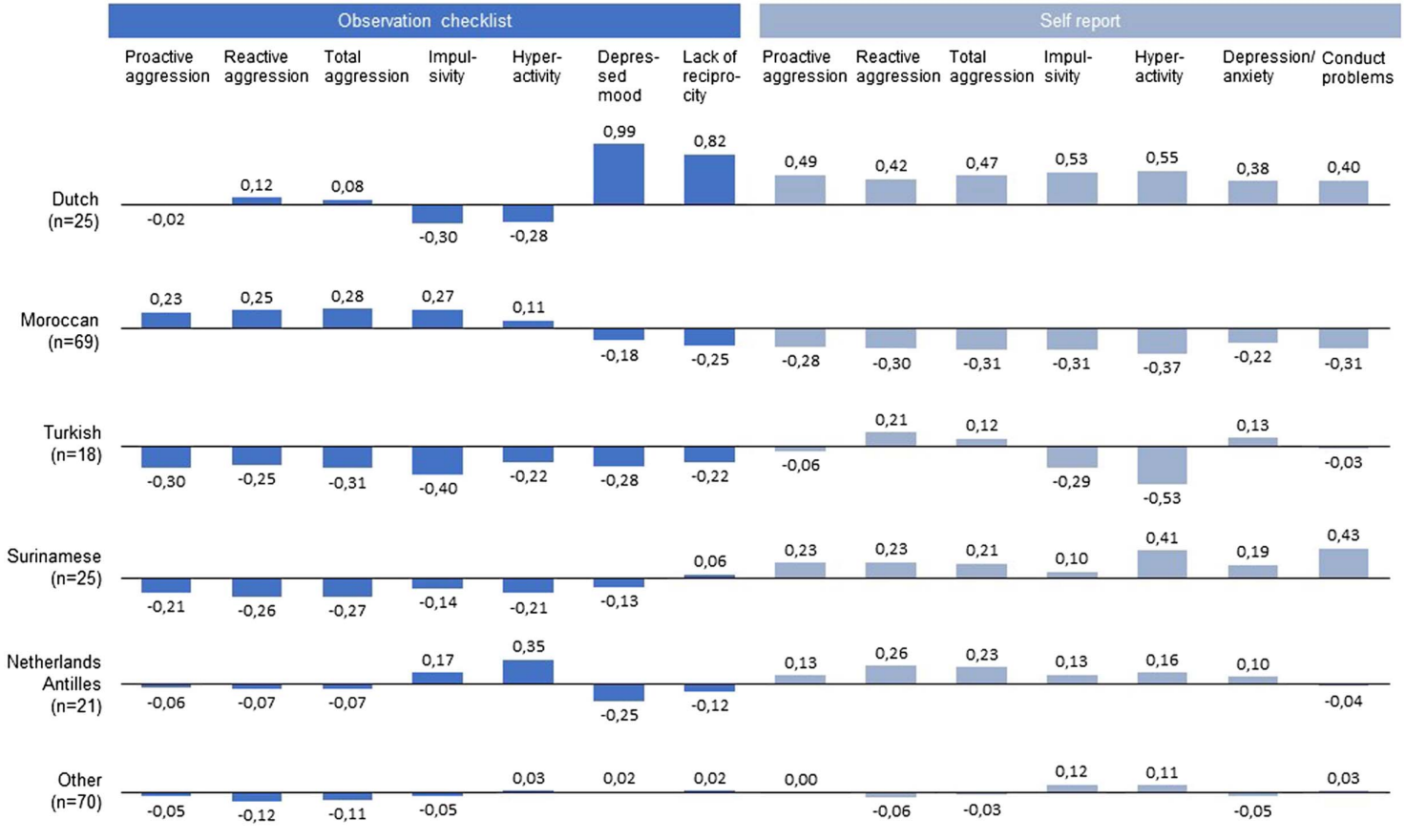
Ethnic Group Differences in the Discrepancy Between OC and Self-Report Scales

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 origins. 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 analyzed 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 origin reported less problems (on the self-report) than were observed, whereas for the youths of Dutch origin, this is the reverse: They reported more problems themselves than were observed and rated on the OC. For proactive aggression, $H(5) = 56.733$, $p = .003$, reactive aggression, $H(5) = 51.326$, $p = .012$, impulsivity, $H(5) = 68.381$, $p = .000$, hyperactivity, $H(5) = 62.490$, $p = .001$, total aggression, SDQ conduct problems, $H(5) = 46.305$, $p = .038$, and total aggression, RPQ total $H(5) = 59.448$, $p = .002$, youths with Moroccan origin differed from youths of Dutch origin. On all the same scales except Impulsivity, the youths with Moroccan origin also differed from the youths with a Surinamese origin with the following values for proactive aggression $H(5) = -52.720$, $p = .007$, reactive aggression $H(5) = -61.288$, $p = .010$, hyperactivity $H(5) = -59.794$, $p = .001$, total aggression (SDQ conduct problems) $H(5) = -45.745$, $p = .043$ and total aggression (RPQ total) $H(5) = -62.997$, $p = .001$. The mixed, other group also differed from the youths with a Moroccan origin on reactive aggression, $H(5) = 33.711$, $p = .037$, impulsivity, $H(5) = -35.900$, $p = .020$, and total aggression, RPQ total $H(5) = 36.168$, $p = .017$. Youth of Moroccan origin differed from youths of Turkish origin on reactive aggression, $H(5) = -57.607$, $p = .014$, and

Figure 1

Z-Scores of Means of OC Self-Report Scales per Ethnic Group



Note. OC = Observation Checklist. See the online article for the color version of this figure.

Table 1
Ethnic Group Comparison on OC and on Self-Report Scales

Observation checklist data	Adj $p < .1$ (2-sided, Bonferroni corrected)	$H(df = 5)$	Adj. p	Self-report data	Adj $p < .1$ (2-sided, Bonferroni corrected)	$H(df = 5)$	Adj. p
Proactive aggression	None			Proactive aggression (RPQ)	$D > M$	8.204	.063
Reactive aggression	$M > O$	10.937	.014	Reactive Aggression (RPQ)	None		
Impulsivity	$M > T$	7.850	.076	Impulsivity (YPI)	$D > M$	8.662	.049
Hyperactivity	$M > O$	8.818	.045	Hyperactivity (SDQ)	$D > M$	10.753	.016
					$D > T$	8.777	.046
					$S > T$	7.351	.049
					$S > M$	9.031	.023
Signs of depressed mood	$D > M$	9.022	.400	Depression/anxiety (MAYSI)	$S > M$	17.897	.000
					$O > M$	7.381	.099
Lack of reciprocity in contact	None						
Aggression composite score	$M > O$	36.408	.007	Conduct problems (SDQ)	None		
				Total Aggression (RPQ)	$D > M$		

Note. Pairwise compared by Kruskal–Wallis, using Dunn–Bonferroni correction for multiple tests. 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; OC = Observation Checklist; RPQ = Reactive Proactive Aggression Questionnaire; YPI = Youth Psychopathic Traits Inventory; SDQ = Strengths and Difficulties Questionnaire; MAYSI = Massachusetts Youth Screening Instrument.

on total aggression, RPQ total, $H(5) = -60.768$, $p = .007$. Juveniles with Turkish origin differed from youths of Dutch origins on hyperactivity, $H(5) = 60.206$, $p = .046$.

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

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 origin 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.

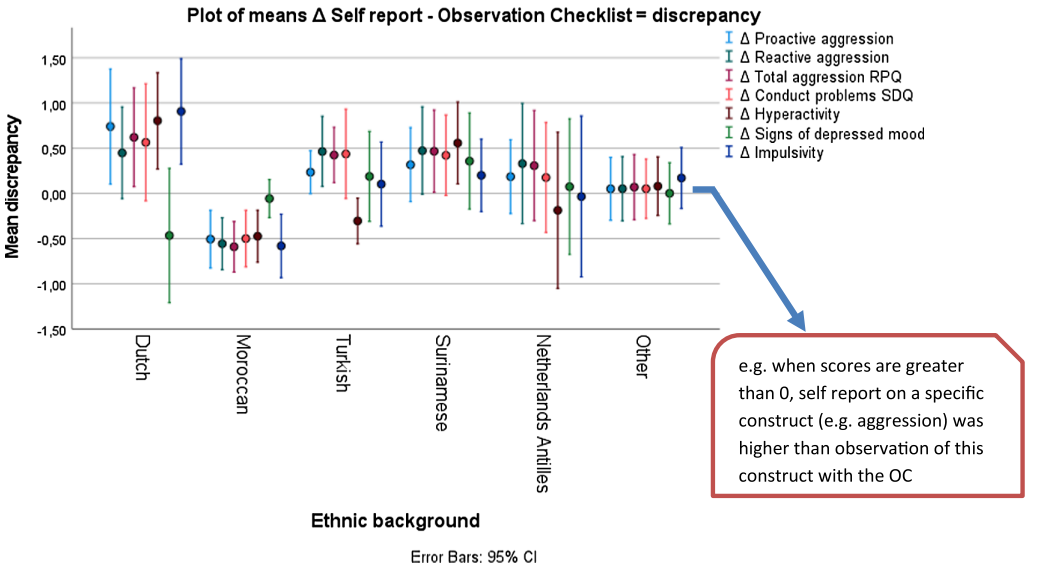
First, between ethnic-group differences were found in the self-report data. Youths of Moroccan origin self-reported fewer problems than Youths of Dutch origins on the more externalizing scales: Aggression, Impulsivity, and Hyperactivity. On Depression/anxiety, they also showed the lowest mean scores, but only significantly different (lower) than the Surinamese youths. On Hyperactivity, the juveniles with Dutch origin reported higher scores than those with Turkish origin as well. Youths with Surinamese origin 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 origins, more than other ethnic groups did. Here, a clear trend was visible: externalizing concepts, for example, total aggression and reactive aggression, were observed the most in youths of Moroccan origin. In contrast, internalizing behavior, for example, signs of depressed mood was observed and rated higher in the youths of Dutch origin 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

Figure 2

Means of the Subtraction of Z-Scores of Self-Report Minus OC



Note. OC = Observation Checklist; RPQ = Reactive Proactive Aggression Questionnaire; SDQ = Strengths and Difficulties Questionnaire; CI = confidence interval. See the online article for the color version of this figure.

origin diverge in the relation between self-report and observation. Youths from Moroccan origin report less problems than are observed, whereas youths of Dutch origin 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 youths of Dutch origin report themselves but no differences between ethnicities in the discrepancies were found.

The first two results, those that show how youths from different ethnic origins 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 et al. (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 that, according to the data, did take place. The lower scoring pattern of youths of Moroccan origin 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 & McKelvey, 1998) and youths of Moroccan origin in the Netherlands (Colins, 2016; Veen et al., 2011). Possible explanations are cultural differences in how psychopathology is defined, different social cultural expectations, shame, not want to be perceived as troublemakers, or fear of judicial consequences (Crone et al., 2010; Davies & McKelvey, 1998; van Batenburg-Eddes et al., 2012). Considering the latter, in the Netherlands, people of Moroccan origin 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 youth of Dutch origin held more often diminished responsible although mental disorders were found similar frequent, and, after pretrial evaluation, less often advised to be placed in a JJI (Vinkers & Duits, 2011). The latter explanation is also given by Veen et al. (2011), who mention disparities in sentencing in the Dutch court system, leading to the incarceration of relatively less troubled youth of Moroccan origin, compared to their peers of Dutch origin. They imply that the lower scores on self-report are reliable, meaning youths of Moroccan origin

indeed have fewer mental health problems. However, other researchers have contested this finding (Vinkers & Duits, 2011). Remarkable is how youths of Surinamese and to a lesser extent those of Netherlands Antilles origin have similar scoring tendencies as the youths of Dutch origin. The shared history of these cultures, as both were colonized by the Netherlands in the 17th century, is perhaps an explanation for more similarities in a foremost language, but also awareness, exposure and expressing of complaints or feelings. The ethnic Turkish youths have more similarities in history and religion with the juveniles with Moroccan origin, often children of (grand)parents that were recruited to work in the Netherlands in the 1960's, at first often with the intention to return (Van Meeteren et al., 2013). Earlier research has found the similar clusters in answering tendencies when researching self-reported crime, for example, youth of Turkish and Moroccan origin self-reporting less police contacts, and those of Dutch, Netherlands Antillean and Surinamese origin more (Junger, 1989; van Batenburg-Eddes et al., 2012). Scoring profiles of youth of Turkish and Moroccan origin are similar on the self-report of hyperactivity, but do differ on the other aspects.

On the observation checklist, group workers rated the juveniles in the group. While possible that youths of Moroccan origin showed more aggression (and thus it was rated more often), there are a number of other possibilities that are highly plausible. First, not registered but noticed, most group workers were of Dutch origin. Research shows that ethnic origin of the subject influences perception of emotional expressions, for example, people from a certain ethnic origin recognize faces and expressions more adequate in people from the same origin, than in cross ethnic groups (Bijlstra et al., 2010; Lipp et al., 2015). Perhaps the elevated scores of depressed mood, as are observed in the youths of Dutch origin but barely in the other groups, can be understood in this direction. Registering origin from the observers, would also be a good starting point for further research. Second, racial or ethnic stereotypes and prejudices have to be considered when interpreting our results. It is interesting how externalizing behavior was observed more in the youths of Moroccan origin, both in contrast with other groups and in contrast with internalizing problems. Future research should aim at detangling this issue, for example, by combining observation and self-report data with institutional incidents or recidivism data. Previous research in line with our findings found that teachers rated internalizing behavior less often in youths of Moroccan origin than boys of Dutch origin, and externalizing behavior more often (Vollebergh et al., 2005). Third, almost all interracial encounters are prone to microaggressions, very possibly leading to different dynamics between subjects and observer, leading to different behavior (Sue et al., 2007). Fourth, it is commonly accepted that how internalizing disorders are expressed or presented differs among cultures (Kirmayer, 2001). The explanation that youths of an immigrant origin can use somatic complaints as an expression of mental troubles (Bengi-Arslan et al., 2002), or use more externalizing gestures to express distress or (agitated) depression, should be taken into account. 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 origins and internalizing and externalizing scales. Differences converge more regarding internalizing problems, but diverge more concerning the more externalizing scales. It becomes clear that youth of Moroccan origins 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 to use many sources of cultural sensitive information. Taking the aforementioned literature together, we believe that 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 "unfavorable" 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 JJI. Despite all these limitations, we are convinced a multifaceted approach in (risk) assessment is key. In concordance with others (; Hunsley & Mash, 2007), we believe that the collection of data from multiple sources should be best practice, as our findings illustrate. Consideration of the meaning of discrepancies both in practice as in research and incorporating cultural sensitivity in this matter is

of urgent importance. Knowledge about the role ethnic origin plays, in subject and observer, and a multimethod assessment yields a more thorough and tailored evaluation.

Some limitations should be addressed: Groups of ethnic origin are relatively small, making it harder to draw conclusions on these groups. However, crucial significant findings emerged, highlighting the power and significance of the reported differences. The contrast of differences with the biggest group, for example, "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 youths of Dutch origin 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 et al., 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 (*article submitted for review*). A final limitation concerns the use of sum scores from the questionnaires used, not taking into account possible different weighing of each item. However, we believe our approach was the most suitable as our groups were relatively small. The weighing of the items would make an interesting point for following research in bigger samples.

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 also be extended to other sources of information, such as parents and teachers next to observation and self-report, concerning this dynamic and complex group. Furthermore, it is of importance to stress that current findings only concern males as female crime numbers are increasing but cultural dynamics can also differ between genders (Leiber & Peck, 2015), more research into the role of gender and culture is essential.

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), 213–232. https://doi.org/10.1037/0033-2909.101.2.213

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]. 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.

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. https://doi.org/10.1007/s001270200003

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. https://doi.org/10.1016/j.jesp.2010.03.006

Bishop, D. M., & Frazier, C. E. (1996). Race effects in juvenile justice decision-making: Findings of statewide analysis. *The Journal of Criminal Law & Criminology*, *86*(2), 392–414. https://doi.org/10.2307/1144031

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. https://doi.org/10.1002/ab.21458

Colins, O. F, 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. https://doi.org/10.1177/070674371005500409

Colins, O. F. (2016). Assessing reactive and proactive aggression in detained adolescents outside of a research context. *Child Psychiatry and Human Development*, *47*(1), 159–172. https://doi.org/10.1007/s10578-015-0553-z

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. https://doi.org/10.1177/1541204015619659

Colins, O. F., Grisso, T., Vahl, P., Guy, L., Mulder, E., Hornby, N., Pronk, C., Markus, M., Doreleijers, T., & 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. https://doi.org/10.1007/s10862-014-9476-4

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.e1. https://doi.org/10.1016/j.jpeds.2009.08.008

Davies, L. C., & McKelvey, R. S. (1998). Emotional and behavioural problems and competencies among immigrant and non-immigrant adolescents. *The Australian and New Zealand Journal of Psychiatry*, *32*(5), 658–665. https://doi.org/10.3109/00048679809113120

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. https://doi.org/10.1080/13557858.2011.645150

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, and Allied Disciplines*, *53*(5), 558–574. https://doi.org/10.1111/j.1469-7610.2012.02537.x

Fazel, S., Doll, H., & Långström, N. (2008). Mental disorders among adolescents in juvenile detention and correctional facilities: A systematic review and metaregression analysis of 25 surveys. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(9), 1010–1019. https://doi.org/10.1097/CHI.0b013e31817eecf3

Ferdinand, R. F., van der Ende, J., & Verhulst, F. C. (2004). Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology*, *113*(2), 198–206. https://doi.org/10.1037/0021-843X.113.2.198

Fite, P. J., Raine, A., Stouthamer-Loeber, M., Loeber, R., & Pardini, D. A. (2009). Reactive and proactive aggression in adolescent males: Examining differential outcomes 10 years later in early adulthood. *Criminal Justice and Behavior*, *37*(2), 141–157. https://doi.org/10.1177/0093854809353051

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. https://doi.org/10.1207/S15374424jccp2901_10

Goodman, R. (1997). The strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

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

Hintze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review*, *34*(4), 507–519. https://doi.org/10.1080/02796015.2005.12088012

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, *3*(1), 29–51. https://doi.org/10.1146/annurev.clinpsy.3.022806.091419

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*, *39*(2), 133–140. https://doi.org/10.1007/s00127-004-0712-1

Junger, M. (1989). Discrepancies between police and self-report data for Dutch racial minorities. *British Journal of Criminology*, *29*(3), 273–284. https://doi.org/10.1093/oxfordjournals.bjc.a047835

Kirmayer, L. J. (2001). Cultural variations in the clinical presentation of depression and anxiety: Implications for diagnosis and treatment. *The Journal of Clinical Psychiatry*, *62*(Suppl. 13), 22–28.

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), 74–96. https://doi.org/10.1037/1040-3590.14.1.74

Lampe, K. G., Mulder, E. A., Colins, O. F., & Vermeiren, R. R. J. M. (in press). An observation checklist for use by residential social workers in juvenile justice institutions. *Journal of Social Work*.

Leiber, M. J., & Peck, J. H. (2015). Race, gender, crime severity, and decision making in the juvenile justice system. *Crime and Delinquency*, *61*(6), 771–797. https://doi.org/10.1177/0011128712446898

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 & Personality Science, *6*(1), 109–115. https://doi.org/10.1177/1948550614546047

Markus, M., Colins, O., Vahl, P., Matser, D., & Vermeiren, R. (2009). *Massachusetts youth screening instrument-second version- Gauthoriseerde Nederlandse vertaling*.

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), 63–72. https://doi.org/10.1093/geronb/52B.2.P63

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. https://doi.org/10.1037/pas0000816

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. https://doi.org/10.1037/h0088258

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, *32*(2), 159–171. https://doi.org/10.1002/ab.20115

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), 31–45. https://doi.org/10.1037/a0033972

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. https://doi.org/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. https://doi.org/10.1016/j.ijlp.2011.08.008

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. https://doi.org/10.1080/13557850802509206

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, *62*(4), 271–286. https://doi.org/10.1037/0003-066X.62.4.271

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. https://doi.org/10.1007/s00787-013-0497-4

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. https://doi.org/10.1177/1477370811421644

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. In J. Ho (Ed.), *Immigrants: Acculturation, socio-economic challenges and cultural psychology* (Immigration in the 21st century: Political, social and economic issues, pp. 113–170). Nova Science Publishers, Inc.

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(12), 1304–1312. https://doi.org/10.1111/j.1469-7610.2008.01942.x

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 Child & Adolescent Psychiatry*, *12*(6), 281–289. https://doi.org/10.1007/s00787-003-0341-3

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. https://doi.org/10.1016/j.ijlp.2011.02.007

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. https://doi.org/10.1016/j.ijlp.2011.02.008

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. https://doi.org/10.1007/s00127-005-0906-1

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. https://doi.org/10.1080/02796015.2005.12088009

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*, *68*(6), 1038–1050. https://doi.org/10.1037/0022-006X.68.6.1038

Received January 28, 2022
Revision received November 1, 2022
Accepted January 1, 2023 ■