



UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Moral development and juvenile sex offending

van Vugt, E.S.

Publication date
2011

[Link to publication](#)

Citation for published version (APA):

van Vugt, E. S. (2011). *Moral development and juvenile sex offending*. Boxpress.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 2: Moral Development and Recidivism: A Meta-Analysis³

3 Van Vugt, E.S., Gibbs, J.C., Stams, G.J.J.M., Bijleveld, C., Van der Laan, P.H., & Hendriks, J. *Moral development and recidivism: A meta-analysis*. *International Journal of Offender Therapy and Comparative Criminology* (2011), Advance online publication. doi:10.1177/0306624X10396441.

Abstract

A meta-analysis of 19 studies ($N = 15,992$ offenders) showed a significant inverse relation between more mature moral development and recidivism. Moderator analyses revealed a larger effect size for moral cognition ($r = .20$) than for moral emotion ($r = .11$). Effect sizes for production measures ($r = .57$) were much larger than for recognition measures ($r = .16$) and unstructured (clinical) judgment ($r = .10$). Larger effect sizes were found for female delinquents ($r = .32$) than for male delinquents ($r = .21$). Only small differences in effect-sizes were found between juvenile delinquents ($r = .10$) and adult delinquents ($r = .16$). Finally, self-report measures of recidivism revealed much larger effect sizes ($r = .32$) than official reports of recidivism ($r = .09$). The discussion focuses on the theoretical and practical meaning of the magnitude of the effect size for the relation between moral development and recidivism.

Moral Development and Recidivism: A Meta-Analysis

Delays or deficiencies in moral development could be meaningful to the accountability of delinquents as well their risk of reoffending. Relevant to the accountability question may be whether offenders evidence a delayed, immature or superficial understanding of right and wrong (moral judgment), and whether they evidence deficiencies in their experience of moral emotions (affective empathy, guilt, or shame) or understanding of emotional states in others (cognitive empathy) (Le Sage, 2005); at least in the judicial systems of some countries, such factors could mitigate accountability. Furthermore, it is important to study the possible relation between moral developmental factors and risk of recidivism. Such information could importantly inform efforts to match the intensity and other qualities of a given intervention program with the risk level of the offender (Andrews, Bonta, & Hoge, 1990; Andrews & Dowden, 1999, 2006; Lowenkamp & Latessa, 2005). Accordingly, we focus in this study on the following questions. How robust are reported inverse relations between moral development and recidivism? Are the relations substantial enough to contribute to risk assessments and selection of intervention programs?

Moral development is a broad concept that includes cognitive and emotional constructs; among these, moral judgment, empathy, guilt and shame have been prominent in the literature. A well established view of moral development and its relation to offending behavior is based on Kohlberg's cognitive developmental approach to moral judgment (Palmer, 2003). In Kohlberg's stage-oriented approach, lower stage moral judgment is dominated by external consequences, such as avoidance of punishment and concrete pragmatic or hedonistic considerations. Higher stage moral judgment, on the other hand, is characterized by reasoning that involves relations with others in which ideal reciprocity, mutual respect, trust and the social contract are emphasized (Gibbs, 2010; Kohlberg, 1984). It is assumed that higher stage moral judgment buffers against antisocial and delinquent behavior, because the well-being of relationships and society is taken into account (Gibbs, 2010; Kohlberg, 1984). In a comprehensive meta-analysis of moral judgment and juvenile delinquency, Stams et al. (2006) found a significant and large association between lower stage moral judgment and juvenile delinquency, even after controlling for socioeconomic status, cultural background, age, intelligence, gender and type of offense.

Moral development is not only conceptualized in terms of moral judgment but also in terms of its emotional facets such as empathy, shame, and guilt. Gibbs (2010) argued that emotional predispositions such as empathy are just as fundamental to moral development, motivation, and behavior as are Kohlbergian cognitive stage structures. Cohen and Strayer (1996) conceptualized empathy as encompassing cognitive and affective components: Empathy's cognitive component marks a

person's understanding of the actual or previous emotional states of others; the affective component indicates the person's ability to share others' emotional states. The ability to empathize is assumed to suppress antisocial, aggressive and other acting out behavior that is harmful (Tangney, Stuewig & Mashek, 2007). Jolliffe and Farrington (2004) conducted a meta-analysis of empathy and offending and showed cognitive empathy to be more strongly and negatively related to offending than affective empathy. Relations between empathy and offending were found to depend on age and type of offense (with larger effect sizes for adolescents compared to adult offenders and smaller effect sizes for sex offenders).

A narrative review by Tangney et al. (2007) discussed the conceptualization of guilt and shame. Although guilt and shame are both moral emotions, their origin seems to differ. Shame could be seen as a more public matter in which a person violates to some degree a society's social and cultural values. In contrast, guilt may be considered a private matter insofar as one reacts to a violation of one's own moral values (Tangney, Miller, Flicker & Hill-Barlow, 1996). Other researchers (e.g., Bradshaw, 1988; Lewis, 1971) have focused not on the situation (private-public) but on the role of the self within the experience. An individual who feels guilty is concerned about the wrongdoing itself and accordingly may seek to undo the violation with a restorative action. In contrast, an individual who feels ashamed is embarrassed and humiliated and concerned about one's own role within the experience resulting in avoidance of the situation that reflects the unpleasant experience. Consequently, shame is considered to be more devastating, in the sense that it focuses on one's self-concept and not on the incorrect behavior. In addition, Tibbets (2003) showed guilt - but not shame - to be negatively related to self-reported delinquency. Some researchers, moreover, even claim shame to be conducive to offending (Hosser, Windzio, & Greve, 2008).

Insofar as cognitive (moral judgment) and affective moral development (empathy, guilt and shame) have been shown to be associated with delinquency (Jolliffe & Farrington, 2004; Stams et al., 2006), one would expect moral development to be also associated with recidivism. The aim of the current study therefore is to add to the literature by examining this relation in order to understand its importance for increased risk for recidivism as well for effectiveness of treatments that aim to prevent delinquents from recidivating. Although several meta-analytic studies of criminal offense recidivism have been conducted to examine potential predictors of recidivism, none of these meta-analyses included studies that predict recidivism from delays or deficiencies in moral development (Bonta, Law & Hanson, 1998; Cottle, Lee & Heilburn, 2001).

This meta-analysis examines the relation between moral development, in terms of cognitive and emotional aspects, and recidivism. Because there are no studies that specifically aim at examining the relation between moral development

and recidivism, we included studies that were not explicitly designed for this purpose, but do report on a relation between moral development and recidivism. Examples of studies that are not designed to examine the tested relation, but from which we were able to extract statistical information, are studies that evaluate risk-assessment tools and interventions. The present study should therefore be considered a first systematic inquiry into the association between moral development and criminal offense recidivism. Meta-analytic studies are valuable, as they allow accurate evaluation of the researched area by analyzing and testing both the strength and direction of relations between constructs, providing opportunities to formulate new hypotheses that cannot be tested in any of the primary studies on which the meta-analysis is based (see Lipsey & Wilson, 2001; McCartney & Rosenthal, 2000).

Accordingly, this study uses meta-analytic techniques to evaluate the effect of moral development on recidivism. We also identify and study factors that might moderate the effect, such as type of moral development (moral emotion or moral cognition), type of instrument (production, recognition measure, or unstructured [clinical] judgment), type of study (cross-sectional vs. prospective study), type of recidivism report, publication status, year of publication, gender, and age.

Method

The present meta-analysis examines the degree to which moral judgment, empathy, guilt, and shame predict recidivism. To be included in the meta-analysis, each study had to (1) examine relations between moral development and recidivism (officially reported criminal offense recidivism or self-report of delinquency after arrest or conviction), and (2) focus within moral development on one or more of three referents. Moral development could refer to (2a) moral judgment in terms of justifying prescriptive social decisions or values by appeals to justice or fairness or related considerations of right and wrong (Gibbs, 2010; Kohlberg, 1958; Rest, 1975); (2b) empathy (cognitive or affective, i.e., the ability to understand or share another's emotional state) (Cohen & Strayer, 1996); or (2c) shame or guilt, two moral emotions that appear after a person attributes an incident to either the self (shame) or to a behavioural act (guilt) (Lewis, 1971) or when one's own (guilt) or others' values (shame) are violated (Tangney et al., 1996).

Multiple search methods have been used in order to avoid biased retrieval of studies published in the major journals, which may selectively publish only the results characterized by lower *p* values and larger effect sizes (Rosenthal, 1995). First, we conducted a computerized search of all relevant databases: PsycLIT, PsycInfo, ERIC, Medline, Psychological Abstracts, National Criminal Justice Reference Service, Cambridge Scientific Criminal Justice Abstracts databases, Dissertation Abstracts and

Google Scholar. No specific year of publication was indicated. The following key words, in varying combinations, were used for our search: moral*, moral judgment, moral reasoning, delinq*, (victim) empathy, guilt, shame, crime, criminal, offend*, offense, re-offense, relapse, recidivism. Second, reference lists from relevant reviews and meta-analysis were used, such as Blasi (1980), Bonta et al. (1998), Bradshaw and Roseborough (2005), Hanson and Morton-Bourgon (2005), Jolliffe and Farrington (2004), Schwalbe (2007), Stams et al. (2006), and Wilson, Bouffard, and Mackenzie (2005). The third step included a search in reference sections of those studies that were drawn from the databases to identify citations that did not appear so far. Last, to overcome the file-drawer problem (Rosenthal, 1991), authors in the field of moral development were contacted and were asked about possible recent studies, unpublished studies, doctoral dissertations, theses, and studies that did not show significant results.

The first and third author of this article coded the moderators that were distinguished for this meta-analysis independently, with a concordance of one hundred percent for both categorical and continuous variables. Beside moderators like gender and age of the delinquent group, we also coded measurement characteristics such as the type of instrument that was used to examine moral judgment, empathy, guilt or shame. We distinguished between production measures, assessing self-produced arguments or descriptions of emotional states by means of open questions, recognition measures with closed questions and unstructured (clinical) judgment. Furthermore, a distinction was made between moral cognition and moral emotion. We coded for type of study, distinguishing between cross-sectional studies comparing moral development of first offenders and repeat offenders, and longitudinal studies examining the relation between moral development and recidivism prospectively. For measuring recidivism, we distinguished between officially reported criminal offense recidivism and self-report of delinquency after a previous arrest or conviction. Finally, we coded publication status, distinguishing between published and unpublished studies (theses, doctoral dissertations and manuscripts) as well as year of publication.

Data Analysis

All statistics were transformed into the effect size r , the correlation between an independent variable (moral development) and dependent variable (recidivism), using Wilson's effect size determination program (2001) and formulas provided by Lipsey and Wilson (2001). When a study did not report the association between moral development and recidivism, but only the nonsignificance of the association, an effect size of zero was assigned. This is a commonly used but conservative strategy, which generally underestimates the true magnitude of effect sizes (Durlak & Lipsey, 1991). Assigning an effect size of zero is preferred to exclusion of the nonsignificant results

from the meta-analysis, as this would result in an overestimation of the magnitude of combined effect sizes (Rosenthal, 1995).

To assess the impact of moral development on recidivism, SPSS macros (Lipsey & Wilson, 2000) were utilized, and both fixed and random effect sizes were computed. The difference between fixed and random effect models concerns the way significance testing is executed. Significance testing in fixed effect models is based on the total number of participants, allowing greater statistical power, but limited generalizability. Significance testing in random effect models is based on the total number of studies included in the meta-analysis, resulting in lower statistical power, but greater generalizability (Rosenthal, 1995).

Homogeneity was tested in order to establish whether the individual study effect sizes are estimating the same population mean, that is, to detect to what extent effect sizes were constant across studies. In case of heterogeneity, there are differences among effect sizes that have some source other than subject-level sampling error, and the overall effect size is not a good descriptor of the distribution of individual study effect sizes. There are real between-study differences that may be associated with different study characteristics (Lipsey & Wilson, 2001, pp. 115-119). When the hypothesis of homogeneity was rejected, moderators were tested to help explain heterogeneity among the effect sizes.

Studies that report significant results are more often accepted for publication than studies that do not report significant results and are therefore less easy to be found. This so called publication bias could result into a file drawer problem, which suggests the sample of studies found for the researched area to be incomplete and not representative for the total sample of studies. To examine whether such publication bias or file drawer problem exists, we calculated the fail-safe number to estimate the number of unpublished studies that were not included in the meta-analysis but in case of inclusion could render the overall significant effect size nonsignificant (Durlak & Lipsey, 1991). Meta-analytic findings are considered to be robust if the fail-safe number exceeds the critical value obtained with Rosenthal's (1995) formula of $5 * k + 10$ in which k is the number of studies used in the meta-analysis. If the fail-safe number falls below this critical value, a publication bias or file drawer problem may exist.

Results

This meta-analysis of the relation between moral development and recidivism consisted of 19 studies reporting on $N = 15,992$ individuals. Table 1 shows an overview of all studies with effect sizes. For the interpretations of the magnitude of effect sizes, the criteria formulated by Cohen (1988) were used. Effect sizes are categorized as $r = .10$ (small), $r = .25$ (moderate) and $r = .40$ (large). In one case, the study by Kantner

(1985), an effect size estimate of $r = 0.00$ was assigned, because this study did not contain sufficient statistical information and reported nonsignificant results.

There did not appear to be an inflationary bias in the results attributable to non-publication of non significant results. Small to medium effect sizes of $r = .11$ ($z = 13.95, p < .001$) and $r = .19$ ($z = 6.57, p < .001$) were found for the relation between moral development and recidivism using the fixed and random effect model, respectively. The fail-safe number for the fixed effect model was $N = 664$, which means that more than 664 studies would need to be found to reduce the overall significant effect size to nonsignificance at $p < .01$. The fail-safe number for the random effect model was $N = 132$. The fail-safe numbers of the fixed and random effect model were both larger than Rosenthal's critical number of 105 ($19 \times 5 + 10 = 105$), suggesting that there was no file drawer effect.

Study	Year	N Del	Sex Del	Juvenile/Adult	Age Del	Instruments	Production, Recognition & Unstructured (clinical) judgment		Type of Study	Registration of Recidivism	Moral Development	r	Status
							Unstructured (clinical) judgment	Production					
Buttel	2002	91	F	A	30.7	Defining Issue Test	Recognition	Recognition	Longitudinal	Official Report	Moral cognition	.32	Published
Barnoski	2004	9692	M/F	J	-	Empathy	Unstructured (clinical) judgment	Unstructured (clinical) judgment	Longitudinal	Official Report	Moral Emotion	.07	Unpublished
Ferwerda, Van Leiden, Arts, & Hauber	2008	824	M/F	J	14.7	Shame	Unstructured (clinical) judgment	Unstructured (clinical) judgment	Longitudinal	Self-Report & Official Report	Moral emotion	.15	Unpublished
Hosser, Windzio, & Greve	2008	157	M	A	21.2	EMO-16-Week	Recognition	Recognition	Longitudinal	Official Report	Moral emotion	.17	Published
Jackson & Bonacker	2006	69	M/F	A	30.7	Mehrabian Emotional Empathy Scale	Recognition	Recognition	Longitudinal	Official Report	Moral emotion	.04	Published
Kendall, Dearthoff, & Finch	1977	67	M	J	15.8	Hogan's Empathy Scale	Recognition	Recognition	Cross-sectional	Official Report	Moral emotion	.03	Published
Kantner	1985	157	M	A	-	Defining Issue Test	Recognition	Recognition	Longitudinal	Official Report	Moral cognition	.00	Published
Lauterbach & Hosser	2007	839	M	A	20.7	Interpersonal Reactivity index	Recognition	Recognition	Longitudinal	Official Report	Moral emotion	.16	Published
Leenman, Gibbs, & Fuller	1983	57	M	J	16.0	Socialmoral Reflection Measure	Production	Production	Longitudinal	Self-report & Official Report	Moral cognition	.75	Published
Little & Robinson	1989	115	M	A	36.6	Defining Issue Test	Recognition	Recognition	Longitudinal	Official Report	Moral cognition	.24	Published
Lodewijks, Doreleijers, de Ruiter, & Borum	2008	66	M	J	15.4	Empathy and Remorse	Unstructured (clinical) judgment	Unstructured (clinical) judgment	Longitudinal	Official Report	Moral emotion	.31	Published
Mityagin	1986	78	M	A	31.2	Socialmoral Reflection Measure & Transgression Guilt Interview	Production	Production	Cross-sectional	Official Report	Moral cognition & emotion	.25	Unpublished
Mulloy, Smiley, & Mawson	1991	68	M	A	36	Interpersonal Reactivity Index	Recognition	Recognition	Longitudinal	Official Report	Moral Emotion	.14	Published
Priest & Kordinak	1991	72	M	A	29.9	Defining Issue Test	Recognition	Recognition	Cross-sectional	Official Report	Moral cognition	.00	Published
Smith & Monastersky	1986	112	M	J	14.1	JSODP	Recognition	Recognition	Longitudinal	Official Report	Moral emotion	.26	Published
Stouthamer-Loeber, Loeber, Wei, Farrington, & Wikström	2002	792	M	J	10.0	Guilt	Unstructured (clinical) judgment	Unstructured (clinical) judgment	Longitudinal	Self Report	Moral emotion	.32	Published
Van der Geest, Bijleveld & Blokland	2007	270	M	J	15.0	Empathy	Unstructured (clinical) judgment	Unstructured (clinical) judgment	Longitudinal	Official Report	Moral emotion	.22	Published
Van der Put	2008	1396	M/F	J	15.0	Moral Conscience	Unstructured (clinical) judgment	Unstructured (clinical) judgment	Longitudinal	Official Report	Moral cognition & emotion	.08	Unpublished
Visser	2004	31	M/F	J	16.0	Moral Orientation Measure	Recognition	Recognition	Longitudinal	Self-Report	Moral cognition & emotion	.44	Unpublished

A homogeneity analysis yielded a significant result, $Q(19) = 109.75, p < .001$, meaning that there was a significant variability in effect sizes between studies. Hence, we conducted categorical and continuous moderator analyses in order to detect possible factors affecting the relation between moral development and recidivism (Mullen, 1989). The categorical moderators were gender, type of instrument (production, recognition measure or [unstructured] clinical judgment), type of moral development, type of study (cross-sectional or longitudinal), type of recidivism report (self-report or official report), and publication status. Age was treated as a categorical variable, juvenile versus adult samples, as it was dichotomously distributed. There was only one continuous moderator, namely year of publication.

We conducted fixed effect and random effect moderator analyses. For only one of the moderators, the random effect model yielded a significant result, which is described in the text below. Table 2 presents an overview of all fixed effect moderator analyses.

Univariate analyses of variance were conducted for the different moderator variables yielding the following results: First, a moderate effect size was found for the relation between moral cognition and recidivism ($r = .20$). The effect size for the relation between moral emotion and recidivism ($r = .11$), $Q_b(1,14) = 3.93, p < .05$, was much smaller. Second, effect sizes were much larger for production measures ($r_{fixed} = .57, r_{random} = .58$) than for recognition measures ($r_{fixed/random} = .16$) and unstructured (clinical) judgment ($r_{fixed} = .10, r_{random} = .17$), both in the fixed and random effect model, $Q_b(2,16) = 37.34, p < .001, Q_b(2,16) = 14.08, p < .001$. Third, the effect size for published studies ($r = .20$) was larger than the effect size for studies that were unpublished ($r = .08$), $Q_b(1,17) = 45.85, p < .001$. Fourth, larger effect sizes were found for female delinquents ($r = .32$) than for male delinquents ($r = .22$) and mixed gender groups ($r = .11$), $Q_b(2,16) = 54.42, p < .001$. Fifth, we found differences in effect sizes for juvenile delinquents ($r = .10$) and adult delinquents ($r = .16$): $Q_b(1,17) = 9.42, p < .01$. Lastly, self report recidivism generated larger effect sizes ($r = .32$) than official report ($r = .09$): $Q_b(1,15) = 41.08, p < .001$.

Table 2: Univariate analysis of variance for moderator variables (fixed effect model)

Moderator variables	N Number of respondents	K Number of studies	Effect size <i>r</i> (fixed effects)	95% confidence interval (fixed effects)	<i>Q</i> statistic between studies	<i>Q</i> statistic within studies
Overall	15,992	19	.11***	.10 to .13		109.65***
Publication status					45.85***	
Published (Journal)	3,971	14	.20***	.17 to .24		45.46***
Unpublished	12,021	5	.08***	.06 to .10		18.34**
Sex					54.42***	
Males only	3,889	13	.22***	.18 to .24		46.58***
Females only	91	1	.32**	.11 to .53		0.00
Mixed	12,012	5	.11***	.06 to .10		8.64
Age (dichotomous)					9.42**	
Juvenile	13,260	10	.10***	.08 to .12		84.61***
Adult	2,732	9	.16***	.13 to .20		15.60*
Conceptualization of moral development ¹					3.93*	
Moral cognition	492	5	.20***	.11 to .29		26.70***
Moral emotion	13,995	11	.11***	.09 to .12		66.54***
Method of assessment					37.34***	
Recognition	2,864	11	.16***	.12 to .20		14.15
Production	135	2	.57***	.40 to .74		3.02
Unstructured (clinical) judgment	12,993	6	.10***	.08 to .11		55.15***

Moderator variables	N Number of respondents	K Number of studies	Effect size <i>r</i> (fixed effects)	95% confidence interval (fixed effects)	<i>Q</i> statistic between studies	<i>Q</i> statistics within studies
Registration of Recidivism ¹						
Self-report	823	2	.32***	.26 to .39	41.08***	0.39
Official Registration	14,288	15	.09***	.08 to .11		44.51***
Type of study						
Cross-sectional	217	3	.17*	.03 to .30		8.72*
Prospective	15,775	16	.11***	.09 to .13		100.24***

* $p < .05$ ** $p < .01$ *** $p < .001$

¹ As the mixed category of this moderator was excluded, the number of respondents does not add up to 15,992.

Discussion

This meta-analysis focused on the relation between moral development and criminal offense recidivism. Small to medium overall effect sizes of $r = .11$ and $r = .19$ were found for the fixed and random effect model, respectively. Moderator analyses revealed differences in effect sizes for moral cognition and moral emotion, with a larger effect size for moral cognition. Effect sizes for production measures were larger than for recognition measures and unstructured (clinical) judgment. Published studies generated larger effect sizes than unpublished studies. Effect sizes were relatively large for female delinquents compared to male delinquents and mixed gender groups. Larger effect sizes were found for adult delinquents than for juvenile delinquents. Finally, self-report recidivism resulted in much larger effect-sizes than official report.

Results from meta-analyses by Stams et al. (2006) and Joliffe and Farrington (2004) indicate that moral cognition is more strongly related to delinquency than moral emotion. The present study is consistent with these meta-analytic results, as a larger effect size was found for the relation between moral cognition and recidivism than for the relation between moral emotion and recidivism. However, Gibbs (2010) argues that moral cognition and moral emotion, although theoretically distinguishable, are intimately interrelated in daily functioning. In addition, Pizarro (2000) contends that although a person without moral emotions could make the same moral judgments as a 'normal' individual, he or she might not be able to recognize a moral situation as this person lacks affective empathy (i.e., sensitivity) to pick up morally relevant cues. Therefore, both moral cognition and moral emotion might be necessary conditions in order to behave morally. Future research should examine whether the integration of moral cognition and moral emotion better predicts recidivism.

Most studies that were included in the meta-analysis examined moral development by means of recognition measures. Recognition measures have been shown to elicit unrealistically high scores in delinquent samples (hence a serious risk of ceiling effects), whereas production measures better assess a proximate of the actual cognitive-affective processes underlying moral motivation, because respondents have to produce arguments or descriptions of emotional states themselves (Stams et al., 2006). It is therefore plausible to suggest that production measures may better predict delinquency and reoffending than recognition measures and unstructured (clinical) judgment. This suggestion is empirically supported by the present study as well as the meta-analysis of moral judgment and delinquency by Stams et al. (2006).

Gender differences were also examined in this meta-analysis. The incidence of delinquency in the male population is much larger than the incidence of delinquency in the female population (Mullis, Cornille, Mullis, & Huber, 2004; Snyder & Sickmund, 1999). Moreover, it is known that women are generally sentenced for less serious

and less violent offenses than men (Acoca, 1999). The somewhat larger effect size for female offenders compared to male offenders in this meta-analysis was based on only one study in which a group of female delinquents was examined who were charged for a relatively serious violent offense, namely, battering. There is empirical evidence showing that females who have entered the justice system suffer from more serious psychopathology than do male delinquents (Hendriks & Bijleveld, 2006; Hendriks & Slotboom, 2007; McCabe, Lansing, Garland, & Hough, 2002). It is possible that the female delinquents of the study that was included in this meta-analysis had serious mental health problems, which may have negatively affected their moral functioning, making them vulnerable for recidivism. Therefore, one should be careful in interpreting the gender effect, here especially since the effect size for mixed groups (men and women) was lower than the effect size for the male group.

A larger effect for the relation between moral development and recidivism was found for adult delinquents than for juvenile delinquents. As adults are typically more advanced in moral development compared to adolescents (Gibbs, Basinger, Grime, & Snarey, 2007), possible ceiling effects in the adolescent group of offenders and greater diversity of scores in the adult group could be responsible for the larger effect size in studies examining adult offenders (Van der Put, 2008). Second, we expect adult delinquents to have longer and more chronic criminal careers, which could also have affected their level of moral development (e.g., resorting to pragmatic, instrumental, or egocentric appeals for justifying the committed criminal acts) resulting in larger effect sizes (Raaijmakers, Engels & Van Hoof, 2005). Altogether, these outcomes support the importance of interventions targeting moral development of adolescent offenders, given the greater prospects for developmental advance in moral functioning as most juvenile delinquents do not yet have a persistent and consolidated or intractable antisocial worldview and lifestyle. Since this meta-analysis showed moral cognition to better predict recidivism, interventions that target moral cognitive processes, such as Equip, might be promising in the reduction of recidivism (Gibbs, Potter, DiBiase, & Devlin, 2009).

The impact of several other variables was also investigated. In general, published studies tend to show somewhat larger effect sizes than unpublished studies (Lipsey & Wilson, 2001), which also proved to be the case for the present meta-analysis. Self-report recidivism showed larger effect sizes for the relation between moral development and recidivism than official report. First, in official crime reports, authorities only report on arrests or convictions. The large number of crimes that remain undetected by the criminal justice system is not represented in official reports, but may appear in self-reports of delinquency, which can therefore generate larger effects. Second, the larger effect size for self-report can be explained by the preservation of reputation hypothesis (Emler & Reicher, 1995), which holds

that juvenile delinquents want to present themselves as “tough” and “unemotional” in a society that is experienced to be hostile to their interests. Such antisocial identity formation in juvenile delinquents might be reflected in both lower levels of moral development and over reporting of delinquent behavior.

We found small to medium overall effect sizes for the relation between moral development and recidivism. One could question, however, what the magnitude of these effect sizes means in the context of recidivism research. A meta-analysis by Schwalbe (2007) showed risk assessment instruments for juvenile justice to predict recidivism with $r = .25$. Gendreau, Little and Goggin (1996) showed adult risk assessment instruments to predict recidivism with a somewhat higher but still moderate effect of $r = .30$. In comparison with these results, the small-to-medium effect sizes for the relation between moral development and recidivism of $r = .11$ (fixed effect model) and $r = .19$ (random effect model) that we found could be considered, if not robust, at least substantial, especially because risk assessment instruments do not predict recidivism from one single factor, but from several risk factors.

Some limitations of this meta-analysis should be mentioned. First, this meta-analysis is based on studies that were not explicitly designed to examine the relation between moral development and recidivism, which mitigate the power to detect moderating effects. Important moderators that could not be tested were the initial offense that led to the first officially registered conviction and the type of recidivism that was reported (e.g., violent offenses, petty crime or sexual offenses). Furthermore, because of the small number of studies reporting on the relation between guilt, shame and recidivism, we were not able to conduct separate moderator analyses for the individual moral emotion constructs.

Second, the results of this meta-analysis might have been affected by restriction of range problems in the level of moral development, as delinquents have been found to show a consistently substantial delay in moral judgment development (Gibbs et al., 2007; Joliffe & Farrington, 2004; Stams et al., 2006). However, there is much heterogeneity to delinquent behavior and the standard deviations, where reported, indicate an adequate range (Gibbs et al., 2007). As noted, lower levels of moral development may create a risk of reoffending, whereas higher stages may protect against adverse environmental influences associated with risk of reoffending (Cohen & Felson, 1979).

Third, the present meta-analysis included several older publications that might have reported with less precision owing to dated methods, using measures that were not designed to predict recidivism from moral development. This might have resulted in an underestimation of the total effect size. To interpret the strength of the effect sizes found in this study, these should be compared with the results of related

studies that theoretically or practically support the importance of the effect sizes (McCartney & Rosenthal, 2000).

Lastly, this study reports on a broad range of reoffending behavior: from shoplifting, assault, dangerous driving, to sex offending. Given such heterogeneity in the offense variable, it is remarkable that effects were found at all. A much stronger effect would possibly be found with a larger set of studies focusing on a more homogeneous set of behavioral outcomes.

We conclude that the inverse relation between moral development and recidivism is substantial and hence relevant to risk assessment, judicial sentencing decisions, and intervention program planning. It is plausible to suggest that the effect size that we found might be an underestimation of the true magnitude of the effect size for the relation between moral development and recidivism, because none of the studies included in this meta-analysis were specifically designed to examine the prediction of recidivism from moral development. Besides, small effects are to be expected when predicting multiple determined behaviors, such as delinquency, from a single predictor (Ahadi & Diener, 1989). Finally, a single factor, measured at a single point in time may underestimate the relation between moral development and recidivism, because both the effects of moral development and delinquency may accumulate over time (see Raaijmakers, Engels & Van Hoof, 2005).

This meta-analysis was a first inquiry into the association between moral development and recidivism showing a stronger effect size for the relation between moral cognition (moral judgment) than for moral emotion and recidivism. Furthermore, this meta-analysis supports the use of production measures for the assessment of moral development, because production measures may better reflect a person's moral performance. Future research is needed to examine whether inclusion of moral developmental constructs such as moral judgment, empathy, guilt and shame, may indeed have incremental value for the prediction of recidivism.