This preprint has been accepted for publication in the *Archives of Sexual Behavior* (06.28.2016)

A META-ANALYSIS OF VIEWING TIME MEASURES OF SEXUAL INTEREST IN CHILDREN

Alexander F. Schmidt[1], Kelly M. Babchishin[2], & Robert J. B. Lehmann[3]

[1]Institute for Health and Behavior, University of Luxembourg.

[2]Royal's Institute of Mental Health Research, University of Ottawa.

[3]Institute of Forensic Psychiatry, Charité – University Medicine Berlin.

Alexander F. Schmidt, Institute for Health and Behavior, Department of Health Promotion and Aggression Prevention, University of Luxembourg, Luxembourg; Kelly M. Babchishin, Royal's Institute of Mental Health Research, University of Ottawa, Ottawa, Ontario, Canada; Robert J. B. Lehmann, Institute of Forensic Psychiatry, Charité – University Medicine Berlin, Germany.

Funding for this project was provided in part by the Canadian Institutes for Health Research Banting Fellowship.

Conflict of Interest: Alexander F. Schmidt is co-author of the Explicit and Implicit Sexual Interest Profile (EISIP; Banse, Schmidt, & Clarbour, 2010) and promotes its non-commercial use for research purposes. Kelly M. Babchishin and Robert J. B. Lehmann declare no conflicts of interest.

Correspondence concerning this article should be addressed to Alexander F. Schmidt, University of Luxembourg, Integrative Research Unit Social and Individual Development

(INSIDE), Health Promotion and Aggression Prevention, Maison des Sciences Humaines, 11, Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg. E-mail: alexander.schmidt@uni.lu

ABSTRACT

Due to unobtrusiveness and ease of implementation, viewing time (VT) measures of sexual interest in children have sparked increasing research interest in forensic contexts over the last two decades. The current study presents two meta-analyses of VT measures adapted to assess pedophilic interest to determine their discrimination between sexual offenders against children (SOC) and non-SOC groups as well as convergent validity (associations with other measures of sexual interest in children). On average, VT measures showed moderate discrimination between criterion groups (fixed-effect d = 0.60, 95% CI [0.51, 0.68], N = 2,705, k = 14) and significant convergent validity with self-reports, penile plethysmography, Implicit Association Tests and offence behavioral measures ranging from r =.18 to r = .38. VT measures, however, provided better discrimination for adults (fixed-effect d = 0.78, 95% CI [0.64, 0.92]) than adolescent samples (fixed-effect d = 0.50, 95% CI [0.40, 0.61]), Q_{between} = 9.37, p = .002. Moreover, using pedophilic difference scores within adult samples substantially increased VT measures' validity (fixed-effect d = 1.03, 95% CI [0.82, 1.25], N = 414, k = 7). Results are discussed in terms of their theoretical and applied implications for forensic contexts.

KEY WORDS: indirect measure, viewing time measure, sexual interest in children, pedophilic interest, meta-analysis, Implicit Association Test

A META-ANALYSIS OF VIEWING TIME MEASURES OF SEXUAL INTEREST IN CHILDREN

INTRODUCTION

Almost 75 years ago, Saul Rosenzweig (1942) introduced the idea that the time spent looking at sexual stimuli could be used as an indicator of sexual interest, with the assumption that sexually appealing stimuli would result in longer latencies. Inspired by the phenomenon that men visiting penny arcades put substantial amounts of money into slot machines to get time-limited visual access to preferred erotic depictions, Rosenzweig created the first indirect latency-based measure of sexual interest that is "…both natural in its imitation of everyday behavior and simple to employ because of its relative freedom from complicated apparatus or interpretative scoring" (Rosenzweig, 1942, p. 150). Rosenzweig presented participants erotic and non-erotic control pictures and allowed participants to choose how long they looked at the pictures while the VTs were unobtrusively recorded. In line with his notion, Rosenzweig (1942) was able to show that groups of schizophrenics differing in sexual activity levels could be discriminated based on VTs for erotic vs. non-erotic stimuli.

Comparing VTs for stimuli of men and women (Zamansky, 1956), Rosenzweig's measure of sexual interest has been adapted into a frequently used, robust latency-based measure of adult sexual orientation (e.g., Ebsworth & Lalumière, 2012; Imhoff et al., 2010; Israel & Strassberg, 2009; Lippa, 2012; Quinsey, Ketsetzis, Earls, & Karamanouikan, 1996; Rönspies et al., 2015). However, not until the end of the 20th century were VT measures validated as a measure of pedophilic sexual interest (Abel, Lawry, Karlstrom, Osborn, & Gillespie, 1994). Since then, a body of research, typically with forensic populations, has been generated that hitherto awaits cumulative integration (for a recent narrative review, see Schmidt, Banse, &

Imhoff, 2015). The current study presents a meta-analysis of VT measures' validity in terms of the ability to discriminate between criterion groups and to converge with conceptually different measures of sexual interest in children.

Psychological Processes Underlying Viewing Time Measures

Over the decades, the general VT effect of sexual orientation has seen multiple independent replications corroborating its robustness. The VT effect emerges when respondents are asked to rate a series of sexually relevant versus irrelevant stimuli in terms of their subjectively perceived sexual attractiveness. Procedural and methodological differences aside, most VT assessments involve asking respondents to rate pictures and, thus, leave the diagnostic purpose of the procedure obvious to the individual being assessed. At the same time, however, the focal dependent variable – rating latency – remains unobtrusive. VT measures can be categorized as *task-relevant indirect measures*[1] due to the fact that the primary task is based on evaluating sexual features. There is preliminary evidence that the primary sexual focus increases group discrimination of task-relevant over task-irrelevant latency-based measures of sexual interest (Rönspies et al., 2015) because the structural overlap between predictor and criterion is maximized (Perugini, Richetin, & Zogmaister, 2010).

What causes longer latencies when sexually attractive versus unattractive stimuli are rated in VT measures? Earlier theoretical reasoning focused on two underlying causal mechanisms: (a) deliberate delay due to the hedonic quality of sexually preferred targets and (b) (automatic) attentional adhesion that slows decision-making after the presentation of explicit erotic material. However, these two causal hypotheses were ruled out in a series of experiments

[1] As opposed to *task-irrelevant indirect measures* wherein the primary task is the classification of non-sexual stimuli features, such as for example color, and sexual stimuli are construed as interfering distractors (see Schmidt et al., 2015, for review).

(Imhoff et al., 2010; Imhoff, Schmidt, Weiß, Young, & Banse, 2012; Schmidt, Imhoff, & Banse, 2014). Specifically, it was shown that prolonged response latencies for sexually relevant versus irrelevant stimuli still emerged when target pictures were removed after a fixed duration prior to the rating task even when possible afterimages were masked (Imhoff et al., 2010, Experiments 1 and 2). Additionally, VT effects occurred when sexual decisions were performed under time pressure (< 1,000ms) or when based exclusively on facial stimuli without any further erotic content (Imhoff et al., 2010, Experiments 3 and 4). Moreover, experimental manipulations of participants' response perspective (e.g., heterosexual men should rate target stimuli from a gay perspective) showed that VT effects were rather a function of the rating perspective than of the stimuli characteristics alone (Imhoff et al., 2012). Finally, without a primary rating task VT effects were absent but could be demonstrated in a sexual attractiveness rating task of completely abstract/non-pictorial symbols representing target age and target gender (Schmidt, Imhoff et al., 2014).

These experiments rule out attentional adhesion and deliberate delay as causal explanations. Instead, findings corroborate that the primary task of scrutinizing a set of criteria subjectively relevant for determining sexual attractiveness is likely causing VT effects of sexual orientation: Denying any single attractiveness criterion results in fast rejection of targets whereas more criteria need to be checked to determine subjective sexual attractiveness which takes longer (see also Pohl, Wolters, & Ponseti, 2015). Therefore, VT effects should better be described as prolonged decision latencies for targets' sexual attractiveness (Imhoff et al., 2010). From an applied perspective the described task-driven effects pose a potential threat to the diagnostic validity of VT paradigms: Only as long as participants comply with the instructions to rate targets' sexual attractiveness, would the measure be expected to produce meaningful results. On the contrary, whenever participants perform sexually irrelevant rating tasks, latency patterns in standard VT paradigms will be invalid (Imhoff et al., 2012; Pohl et al., 2015).

Validity of Viewing Time as a Measure of Sexual Interest in Children

Sexual interest in children is routinely included in risk assessment instruments (e.g., STABLE-2007, Hanson, Harris, Scott, & Helmus, 2007; Sex Offender Risk Appraisal Guide, SORAG, Quinsey, Harris, Rice, & Cormier, 1998) and is an important target in many North American sexual offender treatment programs (e.g., McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). Relatedly, sexual interest in children is one of the best predictors of sexual recidivism among sexual offenders against children ($d = 0.32$, $p < .05$; Hanson & Morton-Bourgon, 2005). Accordingly, the accurate assessment of sexual interest in children is of critical importance in community supervision, risk assessment and management, as well as treatment of sexual offenders. There are several different methodologies available to assess sexual interest in children, each with their strengths and limitations (see Kalmus & Beech, 2005, for review). VT measures are especially appealing to forensic contexts because of their unobtrusiveness.

Validating (not only VT) measures of sexual interest in children is not a trivial task. Currently, there exists no psychometrically flawless criterion for pedophilic interest. Obviously, self-report questionnaires or interviews as direct measures of, for example, sexual fantasies are limited by self-report biases given the severe legal and psychosocial repercussions for individuals who admit such inclinations. Observable behavior as an alternative diagnostic indicator is problematic as the link between sexual behavior involving children and pedophilic interest remains equivocal. Sexual abuse of children is not a regular epiphenomenon of corresponding sexual motivations. On the one hand, only roughly 20% to 50% of convicted child sexual abusers are considered to be pedophilic (Schmidt, Mokros, & Banse, 2013; Seto, 2009, 2010). Among males in the community, on the other hand, many more men self-report indications of sexual interest in children than indicating sexual offences against children (Dombert et al., 2016). Hence, more specific victim characteristics (Screening Scale for Pedophilic Interest SSPI; Seto, & Lalumière, 2001; Seto, Stephens, Lalumière, & Cantor, 2015) and offending behaviors (Dahle, Lehmann, & Richter, 2014; Lehmann, Goodwill, Hanson, & Dahle, 2014) have been associated with sexual interest in children. Phallometric or penile plethysmographic (PPG) assessment of penile tumescence is frequently used as an indirect physiological indicator of sexual arousal although this measure also is not free from methodological and conceptual problems (Kalmus & Beech, 2005; Laws, 2009). Finally, modifications of the Implicit Association Test (IAT; Greenwald, McGee, & Schwartz, 1998) have only recently been reported as valid indirect latency-based measures of sexual interest in children (mean weighted Cohen's d [fixed-effects] = 0.63, 95% CI [0.47, 0.79]; k = 12; N = 707; Babchishin, Nunes, & Herman, 2013). However, at the present stage it remains unclear, how IAT measures of pedophilic interest relate theoretically to other modalities of sexual interest in children such as self-reported fantasies or physiological sexual arousal (Babchishin, Nunes, Hermann, & Malcom, 2015; Schmidt et al., 2015). Ultimately, all of the abovementioned indicators of pedophilic interest have their specific shortcomings. In consequence, validity calculations based on each specific criterion have to be regarded as approximations of the "true" extent of sexual interest in children – independent from the yet open question of an optimal scientific operationalization of the empirically elusive concept of sexual interest in children due to its multimodal nature.

In the literature, the prototypical validation study of VT measures of sexual interest in children uses convicted sexual offenders against children (SOC) as criterion group in comparison against different control groups, such as sexual offenders against adults (SOA), non-sexual offenders (NSO), or non-offenders (NO). Less frequently, VT measures have been used along with external criteria indicative of sexual interest in children such as offending behavioral indexes, self-report, conceptually different indirect latency-based measures, and/or PPG assessments without necessarily implementing control group comparisons (for an descriptive overview of corresponding studies see Table 1).

Methodological Issues: Stimulus Sets and Scoring Algorithms

A crucial factor influencing measurement validity is methodological variability concerning scoring algorithms and procedural aspects, such as stimuli and task characteristics. Similar to PPG research, where lack of standardization is a major criticism (e.g., Kalmus & Beech, 2005; Laws, 2009), one cannot speak of a standardized VT measure but rather of a family of tasks that share the same dependent variable (i.e., decision latencies) but are quite heterogeneous in terms of stimulus characteristics and scoring algorithms.

Stimulus factors. Viewing time measures of sexual interest in children typically rely on comparisons of stimulus sets of child and adult pictures. Although in most countries, 14 or 16 years of age is the legal threshold for sexual contacts, from a psychological point of view this is a problematic demarcation (e.g., Prentky & Barbaree, 2011) as it concerns the possible dissociation between targets' physical age and bodily sexual maturation as a function of pubertal development. Therefore, it is necessary to distinguish between pre-, peri-, and postpubescent developmental stages corresponding to pedophilic, hebephilic, and teleiophilic sexual interests, respectively. Notably, this distinction according to pubescence status must not be mapped onto specific age bands. For the sake of the present study, VT studies vary how stimuli sets are constructed in terms of the differentiation between adults and children according to pubertal stages (i.e., are child categories restricted to prepubescent stimuli or are peripubescent categories included as well). Moreover, stimuli sets differ in degrees of sexual explicitness (nude, partially clothed, or fully clothed stimuli) and type of pictures (computer-generated vs. real photos).

Scoring algorithms. Most VT paradigms use a pedophilic difference index/pedophilic differential to quantify pedophilic sexual interest, which has been shown to maximize validity in PPG research (Blanchard et al., 2009; Harris, Rice, Quinsey, Chaplin, & Earls, 1992). Crucially, this difference value represents a *relative measure* of sexual preference of one target category over the other. Additionally, VT difference measures of one stimulus category over another stimulus category inherently control for potential influences of general processing speed as baseline response latency is canceled out of the final score. Accordingly, general processing speed was not associated with relative VT measures of sexual interest in children (r = -.07; Schmidt, Gykiere et al., 2014). Furthermore, when the highest mean category latency for male or female adult stimuli is subtracted from the highest mean category latency for male or female child stimuli the resulting maximized pedophilic differential score also controls for respondents' sexual gender preferences (i.e., sexual orientation). Finally, using category aggregates increases reliability and prevents capitalizing on idiosyncratic characteristics that influence latencies for single stimuli (i.e., outliers). Only the latter advantage refers also to the use of aggregated VTs for single target categories as an *absolute measure* of sexual interest. Absolute sexual interest levels might be of interest due to the fact that, for example, Harris et al. (1996) have shown that SOC's VT sexual interest levels across all stimulus categories are reduced in comparison with NO. This kind of information will be missed when resorting to relative measures of sexual interest exclusively. In summary, scoring methods might add method variance to the literature and, thus, the scoring algorithm (relative versus absolute) will be tested as a potential moderator in the present meta-analysis.

PRESENT META-ANALYSIS

The current study presents a meta-analysis of the ability of VT measures of sexual interest in children to distinguish between criterion groups and its convergent validity with conceptually different measures of pedophilic interest. In addition to common moderator variables (i.e., published vs. unpublished research, peer reviewed vs. non-peer reviewed studies, publication year), we tested possible moderators of the aggregated effect sizes such as scoring method (relative pedophilic difference indexes vs. absolute VTs), children stimuli type (only prepubescent stimuli vs. mixture), type of pictures (computer-generated vs. real photos) sample location (institution vs. community), type of SOC criterion (index is SOC vs. any history of SOC), adult vs. juvenile SOC samples, treatment status, and type of control group. Because NO controls are much more different from SOC than offender controls and among offending populations NSO should be less similar to SOC than SOA in terms of all kinds of control variables (e.g., range of paraphilic interests, executive functioning, socio-economic status, self-regulation skills, etc.) we hypothesized for the latter moderator the more different the control groups the larger the contrast with the criterion group. Accordingly, we expected a linear trend of effect sizes from unspecific to more specific contrasts based on the similarity of the respective contrast groups (i.e., SOC vs. NO > SOC vs. NSO > SOC vs. SOA; see also Babchishin et al., 2013). Moreover, in line with findings from PPG research (Harris et al., 1992) we expected relative pedophilic difference indexes to produce larger VT effects than scores based on absolute VTs.

METHOD

Selection of Studies

Online searches for studies on VT measures of sexual interest in children were conducted through PsycINFO, ProQuest Dissertations and Theses, Web of Science, and Medline using a combination of search terms: *child* molest** or *child* sex* abuse** or *sex* offend** and (*viewing time* or *reaction time* or *latenc** or *Abel* or *Affinity* or (*implicit* or *indirect*) and (*measure** or *assess**)). Additional studies were identified by reviewing the reference lists of collected studies, conference proceedings from the Association for the Treatment of Sexual Abusers, contacting researchers, and utilizing Google Scholar. The search ended on 07.20.2015 and resulted in 19 eligible studies representing the same number of unique, non-overlapping samples (7 US-American samples, 4 Canadian samples, 1 mixed US-American/Canadian sample, 4 German samples, 2 British samples, 1 Belgian sample). To be included in the current meta-analysis, eligible studies had to report on an identifiable sample of male SOC as well as a comparison sample of male SOA, NSO, or NO that were scored on a VT measure of sexual interest in children (Table 1). If studies included no comparison group, at least VT correlations with measures of external criteria indicative of sexual interest in children (i.e., self-report, PPG assessment, SSPI) or actuarial risk scales had to be reported (Table 1). All studies had to include sufficient statistical information to calculate relevant effect sizes (Cohen's *d* for group comparisons, *r* for correlations with external criteria) and each subsample had to consist at minimum of seven individuals. Unfortunately, due to a lack of sufficient information on relevant effect sizes some potentially informative samples had to be excluded (Abel et al., 1994; Abel et al., 1998; Giotakis, 2005; Gray & Plaud, 2005).

(insert Table 1 about here)

Coding Procedure

Each study was coded with a standard list of variables and explicit coding rules (the coding manual with standard variable list is available upon request from the authors). Each study was coded by the first and third author to conduct interrater analyses and to generate final consensus ratings. Ratings had two components: information describing the study (one form per study) and effect size information (one form per effect size). At the end of coding process, only variables that included data from at least three studies were included in the analyses.

Interrater Reliability. Continuous variables were assessed using absolute intra-class correlations [ICC] based on a two-way mixed design. Cicchetti (1994) suggests interpretive guidelines for ICC ratings of .40 as fair agreement, .60 as good agreement, and .75 as excellent agreement. Categorical variables were assessed using Cohen's κ statistic and percent agreement. Landis and Koch (1977) suggest interpretive guidelines for Cohen's κ of .21 for fair agreement, .41 for moderate agreement, .61 for substantial agreement, and .81 for almost perfect agreement. Interrater reliability analyses for the descriptive statistics and moderators were based on 15 studies and excluded the four studies that were identified after the interrater analyses. Two sets of effect sizes were coded: Cohen's *d* (group discrimination) and correlation coefficients (convergent validity).

Both raters coded a total of effect sizes with high levels of agreement (for Cohen's *d* = absolute intra-class correlation [ICC] based on two-way mixed model and single measure = .881, *n* = 47; for correlations: ICC = .872, *n* = 13). Interrater reliability for continuous variables ranged from ICC = .999 to 1.00 (*Mdn* = 1.00, *n* = 13). For categorical variables, interrater reliability ranged from 69% to 100% agreement (*Mdn* = 92%, *n* = 29; κ ranged from .50 to 1.00, *Mdn* = .85, *n* = 27). Criteria for the SOC group had moderate level of agreement (69% agreement; κ = .50); the remaining categorical variables had substantial to perfect agreement (Landis & Koch, 1977).

None of the variables were excluded due to unacceptable interrater reliability. For all studies, a consensus rating between the two raters was completed after interrater reliability analyses were conducted.

Overview of Analyses

Effect sizes. Two effect size indicators were meta-analyzed: Cohen's *d* and correlation coefficient *r*. The standardized mean difference (Cohen's *d*) is defined as follows: $d = (M_1 - M_2)/S_w$, where M_1 and M_2 are the group means, and S_w is the pooled within standard deviation (Hasselblad & Hedges, 1995). According to Cohen (1988), *d* values of 0.20 are considered small, 0.50, medium, and 0.80 large. A positive *d* indicates that SOC had higher VT scores (indicative of more pedophilic interest) than the comparison group. Importantly, effect sizes based on absolute VTs were aggregated over female and male child categories as only one effect size could be used per sample if data integration adheres to the principle of independent observations. This, however, comes at the price that sexual orientation is not taken into account. In case of maximized relative difference scores this was not a problem and these were used for the meta-analytic integration (see Discussion section for a more detailed elaboration on this issue).

We also meta-analyzed Pearson's *r*, which indexes the direction and size of the relationship between two variables (e.g., the VT measure and another measure of sexual interest in children). Pearson's *r* values of .10 are considered small, .30 moderate, and .50 large (Cohen, 1988). Correlations were transformed into Fisher's *z* and the meta-analysis was conducted on *z*-scores (so that the weight attributed to studies is no longer influenced by the size of the correlations; Borenstein, Hedges, Higgins, & Rothstein, 2009). We converted the meta-analytical findings back to Pearson's *r* for ease of interpretation.

Aggregation of findings. Findings across studies were aggregated using fixed-effect and random-effects meta-analysis (Borenstein et al., 2009). Whereas the results of fixed-effect meta-analysis are conceptually restricted to the particular set of studies included in the meta-analysis, random-effects meta-analysis estimates effects for the population of which the current sample of studies is a part. When variability across studies is low (*Q* < degrees of freedom), random-effects and fixed-effect meta-analysis produce identical results. When the analysis includes a small number of studies (*k* < 30), greater interpretive weight should be given to fixed-effect rather than random-effects analyses because the between-study variability estimate necessary for random-effects analyses loses precision (Schulze, 2007).

To test the variability of findings across studies, we used Cochran's *Q* statistic and the I^2 statistic (Borenstein et al., 2009). The *Q* statistic provides a significance test for variability, whereas the I^2 is an effect size measure for variability and can, therefore, be compared across analyses. As a rough heuristic, I^2 values of 25%, 50%, and 75% can be considered low, moderate, and high variability, respectively (Higgins, Thompson, Deeks, & Altman, 2003).

Following Hanson and Morton-Bourgon (2005), a finding was considered an outlier if it was the single extreme value and accounted for more than 50% of the total variance (*Q*), and the overall variability (*Q*) was significant. When outliers were identified, results are presented both with and without the outlier, with the main interpretation focusing on the findings with the outlier removed. The exception is that if an analysis of three studies identified one study as an outlier, it was not removed (with so few studies, identifying outliers becomes unstable).

Moderator analyses. Fixed-effect meta-regression was used to examine the extent to which the continuous moderator variables influenced the magnitude of group differences whereas the between-level *Q* statistic was used for categorical moderator variables. The overall

Q statistic was partitioned into variability across samples that could be explained by the moderator (between-level variability, which will be referred to as between-level *Q*), and unexplained variability within each level of the moderator (within-level variability, which will be referred to as *Q*). A significant between-level *Q* statistic indicates that the moderator variable explained a significant portion of the variability across samples. The *Q* statistic is distributed as a χ^2, with x - 1 degrees of freedom (x = the number of levels of a moderator).

RESULTS

Group Discrimination Studies

Table 2 presents the meta-analysis of the accuracy of VT scores in discriminating SOC from non-SOC. Overall, VT measures of sexual interest in children were able to distinguish SOC from non-SOC (fixed-effect d = 0.60, 95% CI [0.51, 0.68], N = 2,705, k = 14; see Figure 1). The effect size is moderate in size and there was moderate heterogeneity across studies (I^2 = 66%). Weighted effect sizes were highest (indicating greater differences between groups on VT measures) when the comparison group was comprised of NO (fixed-effect d = 0.84, 95% CI [0.66, 1.03], N = 548, k = 7), followed by NSO (fixed-effect d = 0.57, 95% CI [0.38, 0.77], N = 428, k = 6), and SOA (fixed-effect d = 0.52, 95% CI [0.42, 0.62], N = 2,705, k = 14). The VT measure provided significantly better accuracy in discriminating SOC from NO than SOC from SOA (p < .01, as evidenced by non-overlapping 95% confidence intervals), thus corroborating our hypothesis of a linear trend following increasing specificity of contrast groups.

(insert Table 2 and Figure 1 about here)

Moderators. Table 3 presents the results of the categorical moderator analyses for the comparison group studies. Viewing time measures provided significantly lower discrimination in juvenile samples (fixed-effect d = 0.50, 95% CI [0.40, 0.61], N = 1,782, k = 2) compared to adult samples (fixed-effect d = 0.78, 95% CI [0.64, 0.92], N = 923, k = 12), Q_{between} = 9.37, p = .002. As there were only two juvenile samples that provided group comparisons subsequent moderator analyses were restricted to adult SOC samples (Abel et al., 2004; Worling et al., 2006).

Scoring method was found to be a statistically significant moderator, Q_{between} = 24.81, p < .001. Absolute VT provided lower group discrimination (fixed-effect d = 0.63, 95% CI [0.41, 0.85], N = 396, k = 4) compared to pedophilic difference scores (fixed-effect d = 1.03, 95% CI [0.82, 1.25], N = 414, k = 7). The degree of nudity approached statistical significance (Q_{between} = 5.06, p = .079), with VT measures comprised of nude pictures providing better discrimination than clothed pictures, and VT measures presenting both nude and clothed pictures being in the middle. The type of pictures (computer-generated vs. real photos) also did not moderate the effect ($\text{Q}_{\text{between}}$ = 0.72, p = .390); there was a large amount of variability in the effect sizes of studies that used real pictures (78%; although no statistical outliers were identified). Publication status, peer-reviewed status, and criteria for classifying SOC groups were not statistically significant moderators (see Table 3). Orwin's fail-safe N indicated that 35 null effect studies would be required to bring the average weighted effect size to a trivial effect (defined as d < .20). Although of interest, treatment participation could not be used as a moderator: Only one study sampled SOC who were not treated, the remaining studies were at least partly treated (67%) or had unknown treatment status (17%). Lastly, we also conducted two meta-regressions. We found that publication year (Z = 1.227, p = .220) and mean age of the SOC group (Z = 0.795, p = .427) did not moderate the observed effect sizes.

(insert Table 3 about here)

Convergent Validity with External Criteria

Only a minority of studies examined the relationship between VT measures and other measures of sexual interest in children. As can be seen in Table 4, VT measures were associated with small to moderate effect sizes to self-report, physiological, IAT, and sexual offence history measures of sexual interest in children (see Figure 2).

(insert Table 4 and Figure 2 about here)

Finally, only two studies examined the relationship between VT measures and actuarial (static) as well as dynamic measures of risk for sexual recidivism (Static-99/R vs. STABLE-2000/STABLE-2007/SVR-20 in Babchishin et al., 2013; Schmidt, Gykiere et al., 2014) and, hence, a meta-analysis could not be conducted. Both these studies, however, found statistically significant moderate effect sizes for static (*r* = .33 in both studies) but not for dynamic risk (*r*s ranging from -.07 to .19).

DISCUSSION

In the present meta-analysis we sought to cumulatively integrate empirical findings on the validity of VT measures as indicators of sexual interest in children. We found a moderate sized weighted effect in discriminating SOC from non-SOC control groups (fixed-effect *d* = 0.60, 95% CI [0.51, 0.68], *N* = 2,705, *k* = 14). The magnitude of this effect is similar to the meta-analytic effect reported for IATs of sexual interest in children (fixed-effect *d* = 0.63, 95% CI [0.47, 0.79]; *k* = 12; *N* = 707; Babchishin et al., 2013) – another task-relevant indirect latency-based measure in forensic research (Schmidt et al., 2015). Importantly, we did not find any indication for publication bias (e.g., published, fixed-effect *d* = 0.78 vs. unpublished fixed-effect *d* = 0.76). Moreover, VT group effects were independent of how SOC were operationalized (i.e., SOC based on index offence vs. any history of SOC) and neither mean SOC group age nor publication year were moderators. These findings emphasize the general robustness of VT effects.

Evidence of known-groups discrimination is further corroborated by significant small to moderate weighted correlations with diverse external criteria of sexual interest in children, such as victim characteristics (*r* = .21; SSPI, Seto & Lalumière, 2001), phallometric assessments (*r* = .25), IATs (*r* = .18), and self-reports (*r* = .38). In sum, the current meta-analysis suggests that, on average, VT measures of sexual interest in children produce meaningful differences between criterion groups and are related to conceptually different indicators of pedophilic interest.

Variability of Viewing Time Validity

We identified moderators that can explain the observed variability in VT effects. First, as hypothesized, effect sizes were a direct function of the type of contrast group: The more similar control groups became, the smaller were the corresponding VT effects (Table 2). For example, the mean weighted VT effects for comparisons with NO led to increases of effects sizes by roughly 60% compared to the average meta-analytic VT effect. This could be due to the fact that NO should differ in more aspects (e.g., range of paraphilic sexual interests, cognitive abilities, socio-economic status, etc.) from SOC than other types of sexual offenders (e.g., rapists) and, thus, are likely to produce larger effects.

Second, VT effects were significantly more pronounced in adult than in juvenile SOC (*d* = 0.78 vs. 0.50, respectively). This finding may be explained by a larger fraction of SOC with genuine pedophilic interest in the adult samples as compared to the juvenile samples. According to DSM-5 (American Psychiatric Association, 2013), a gap of at least five years of age between a SOC and his victim is a necessary precondition for the diagnosis of pedophilic interest/disorder. In samples that by definition are not older than 18 years of age, this criterion is more difficult to fulfill. Therefore, juvenile SOC were more likely to offend against minors who might have been subjectively perceived as peers than adult offenders with a much larger age and sexual maturation gap. In line with this notion, sexual victim profiles indicative of sexual interest in children (i.e., SSPI; Seto & Lalumière, 2001) showed less strong associations with phallometrically assessed pedophilic interest in adolescents than in adults (Seto, Murphy, Page, & Ennis, 2003). The reduced pedophilic sexual interest levels of adolescent vs. adult SOC, however, need not necessarily contradict studies that have found that PPG is a valid indicator of sexual interest in children for adolescent sex offenders (Rice, Harris, Lang & Chaplin, 2012; see Ryan, 2016, for review).

Third, restricting the database to adult samples exclusively, the validity of VT measures was substantially increased to a conventionally large effect size when relative pedophilic difference scores were used instead of VT measures based on absolute VTs for child categories (*d* = 1.03 vs. 0.63, respectively). This is concomitant with earlier findings from Harris and colleagues (1992) who showed that maximized difference scores also increased the validity of PPG assessments. Utilizing optimal scoring algorithms, thus, substantially increases mean weighted VT effects by roughly 70% and raises the validity of VT measures of sexual interest in children well above indirect latency-based measurement alternatives, such as the IAT (*d* = 0.63; Babchishin et al., 2013). This has important implications for applied forensic contexts: First of all, if scored properly (i.e., difference scores), VT outperforms IATs in terms of known-groups discrimination. This finding is corroborated in all available studies comparing both VT and IAT paradigms in the same sample (Babchishin et al., 2014; Banse, Schmidt, & Clarbour, 2010; Schmidt, Gykiere, et al., 2014; Schmidt et al., 2013, Schmidt, Bonus, & Banse, 2010). Interestingly, although IAT measures are inherently based on a difference score they still fall short of VT difference scores' validity. Most importantly, maximized difference scores overcome the problem of sexual orientation (i.e., sexually favoring one target sex over the other reduces mean differences when averaging categories across both sexes) and individual differences in cognitive abilities (i.e., due to inherently controlling for confounds by contrasting two comparison categories). Hence, maximized difference scores have to be favored over scores based on absolute category VTs if absolute sexual interest differences are not the prime variable of interest.

Finally, stimulus type – but not whether pictures are computer-generated or represent real photos – may moderate VT effects (marginally statistically significant at p = .079) as effect sizes increased with the degree of nudity in the pictorial stimuli (d = 1.09 in nude vs. 0.68 in fully clothed stimuli; Table 3). Notably, as outlined in the introduction, research has shown that VT effects (at least in non-forensic populations) are driven to a much larger extent by task characteristics (i.e., rating the subjective sexual attractiveness of stimuli) than by stimulus features alone (Imhoff et al., 2012). Although all VT variants in the present research had implemented similar rating tasks, our findings corroborate that stimulus features such as degree of nudity increases effect sizes by roughly two thirds. This is in contrast to findings reported in Schmidt, Imhoff, & Banse (2014) where nude vs. partially clothed stimuli of adults did not interact with the VT effect of sexual orientation in men and women from the community. Hence, this is not only an interesting theoretical research question that warrants further research but is particularly relevant for applied forensic assessment purposes where it is ethically much more adequate to use partially clothed stimuli than pictures involving nude children. At the same time, according to the present findings, using ethically unproblematic pictures comes at the cost of diagnostic accuracy for a construct that is substantially linked to sexual reoffending risk (e.g., Gray et al., 2015; Mann, Hanson, & Thornton, 2010). This poses a vexing dilemma of two ethical goods that stand in opposition to each other.

Limitations and Outlook

The current meta-analysis is limited by the actual number of existing studies that report a sufficient amount of statistical information necessary for meta-analytic integration. In the present case, the number of available independent studies (k = 14 for group comparisons) has to be considered relatively small for a meta-analysis. To deal with this limitation, we based our interpretations exclusively on the results from fixed-effects meta-analyses (Schulze, 2007), although we reported random-effect models in all corresponding tables (without substantial variation in terms of the cumulated effect sizes). For the meta-analysis of convergent validity with external criteria of sexual interest in children the available database was particularly small with independent samples ranging from four to seven studies. Although promising in general, the convergent validity results have to be regarded as preliminary at the present stage.

Nevertheless, the reported mean weighted effect for comparisons of SOC and non-SOC controls (fixed-effect d = 0.60, 95% CI [0.51, 0.68], N = 2,705, k = 14; rising to a substantial d = 1.03 in case of using difference scores) has to be considered as a lower bound of the potential validity of VT measures of sexual interest in children. As outlined in the introduction, the variability of true pedophilic interest in the comparison groups is not fully explained by offending status alone. Not every SOC is necessarily an exclusive or a non-exclusive pedophilic individual (Dombert et al., 2016; Seto, 2009). Thus, for the present results, it has to be kept in mind that the pedophilic interest criterion groups (i.e., SOC) are only proxy groups with an increased but unspecified likelihood of containing a higher fraction of males with sexual interest in children. As such, the outcome of comparisons of SOC vs. non-SOC groups on measures of sexual interest in children will be influenced by the actual fraction of individuals with such a paraphilic interest in these groups. Hence, under ideal conditions of “pure” criterion groups the validity of VT measures will much likely be higher. Therefore, future studies should include additional information on conceptually different proxy measures that help to identify the amount of pedophilic interest at least in the SOC groups (e.g., SSPI levels, boy-girl sexual interest ratios, or PPG difference score levels across comparison groups). These additional sample descriptors could be used as potential moderators in further meta-analyses.

Only two studies examined the cross-sectional link between VT measures and risk of recidivism (Babchishin et al., 2013; Schmidt, Gykiere et al., 2014). Moderate effect sizes were reported for static risk factors (*r* = .33 in both studies) primarily tapping into past criminal and sexually deviant behavior but not for a broader range of fluctuating and potentially changeable dynamic risk indicators indicative of self-regulation problems (*r*s ranging from -.07 to .19). Despite the fact that research shows lower recidivism rates over the last few years (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012), VT measures of sexual interest in children have been prospectively linked to sexual reoffending recently (Gray et al., 2015), further justifying their usefulness for forensic assessments. Of note, associations with risk levels can only be regarded as an indirect proof of construct validity for sexual interest in children as recidivism risk is only partially driven by pedophilic sexual interest (e.g., Brouillete-Alarie, Babchishin, Hanson, & Helmus, 2016; Mann et al., 2010). Finally, for applied forensic purposes it remains an important open empirical question at the present stage whether VT measures are amenable to faking attempts once the measurement rationale is known to the respondents or when process-naïve individuals try their subjective best to manipulate the assessment outcome. It is surprising that the fakeability of VT measures has yet to be empirically researched.

Conclusions

In summary, VT measures of sexual interest in children can be regarded as a valid indirect latency-based measurement and a helpful adjunct to other available measures. At present, VT measures can be considered the best validated indirect latency-based measure of sexual interest in children and, thus, have to be preferred over corresponding IATs. Due to the lack of studies with other external validation criteria and the multimodal nature of pedophilic sexual interest itself, future studies should strive to incorporate conceptually different measures of sexual interest in children for triangulation.

One big advantage of VT measures is the ease of technical implementation and scoring as well as the ease of the instructions. This renders VT measures of sexual interest as a highly flexible tool to investigate all kinds of atypical and/or paraphilic sexual interests (see Larue et al., 2014 for a VT measure tapping into sexual preferences for sexual violence). Nevertheless, we end with a cautionary statement as VT measures rely on their measurement rationale not being transparent to the respondents. Therefore, it is debatable whether it has been a wise decision that the DSM-5 explicitly refers to "*viewing time*" as valid indirect assessment of pedophilic sexual interest (American Psychiatric Association, 2013, p. 699). In short, the simplicity and parsimonious nature of VT tasks that renders them attractive to researchers and diagnosticians alike might at the same time pose a significant danger to their validity.

Compliance with Ethical Standards:

Funding: This project was funded in part by a grant from the Canadian Institutes for Health Research Banting Fellowship to Kelly M. Babchishin.

Ethical Approval: This article is a secondary meta-analysis of data from research that has been published before. Hence, no human participants have been tested for the purpose of the present research.

References

Asterisks (*) indicate studies included in the meta-analysis.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, DSM-5* (5th ed.). Washington, DC: American Psychiatric Association.

Abel, G. G., Huffman, J., Warberg, B., & Holland, C. L. (1998). Visual reaction time and plethysmography as measures of sexual interest in child molesters. *Sexual Abuse: A Journal of Research and Treatment, 10,* 81–95. doi:10.1177/107906329801000202

*Abel, G. G., Jordan, A., Rouleau, J. L., Emerick, R., Barboza-Whitehead, S., & Osborn, C. (2004). Use of visual reaction time to assess male adolescents who molest children. *Sexual Abuse: A Journal of Research and Treatment, 16,* 255–265. doi: 10.1177/107906320401600306

Abel, G. G., Lawry, S. S., Karlstrom, E., Osborn, C. A., Gillespie, C. F. (1994). Screening tests for pedophilia. *Criminal Justice and Behavior, 21,* 115-131. doi: 10.1177/0093854894021001008

Babchishin, K. M., Nunes, K. L., & Hermann, C. (2013). The validity of Implicit Association Test (IAT) measures of sexual attraction to children: A meta-analysis. *Archives of Sexual Behavior, 42,* 489-499. doi:10.1007/s10508-012-0022-8

Babchishin, K. M., Nunes, K. L., Hermann, C. A., & Malcom, J. R. (2015). Implicit sexual interest in children: does separating gender influence discrimination when using the Implicit Association Test? *Journal of Sexual Aggression, 21*, 194-208. doi: 10.1080/13552600.2013.836575

*Babchishin, K. M., Nunes, K. L., & Kessous, N. (2014). A multimodal examination of sexual interest in children. *Sexual Abuse: A Journal of Research and Treatment, 26,* 343-374. doi:10.1177/1079063213492343

*Banse, R., Schmidt, A. F., & Clarbour, J. (2010). Indirect measures of sexual interest in child sex offenders: A multimethod approach. *Criminal Justice and Behavior*, *37*, 319–335. doi:10.1177/0093854809357598

Blanchard, R., Kuban, M. E., Blak, T., Cantor, J. M., Klassen, P. E., & Dickey, R. (2009). Absolute versus relative ascertainment of pedophilia in men. *Sexual Abuse: A Journal of Research and Treatment, 21,* 431-441. doi: 10.1177/1079063209347906

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York, NY: Wiley. doi: 10.1002/9780470743386

Brouillete-Alarie, S., Babchishin, K. M., Hanson, K. R., & Helmus, L.-M. (2016). Latent constructs of the Static-99R and Static-2002R: A three-factor solution. *Assessment, 23,* 96-111.. doi: 10.1177/1073191114568114

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment. 6*, 284–290. doi: 10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge.

Dahle, K.-P., Lehmann, R. J. B., & Richter, A. (2014). Die Screening Skala Pädophilen Tatverhaltens [The screening scale of pedophilic criminal behavior]. *Forensische Psychiatrie, Psychologie, Kriminologie, 8,* 208-215. doi: 10.1007/s11757-014-0261-8

Dombert, B., Schmidt, A. F., Banse, R., Briken, P., Hoyer, J., Neutze, J., & Osterheider, M. (2016). How common is males' self-reported sexual interest in prepubescent children? *Journal of Sex Research, 53,* 214-223. doi: 10.1080/00224499.2015.1020108

*Douroux, A .N. (2013). *A comparison of the Screening Scale for Pedophilic Interest and the Abel Assessment for Sexual Interest-2* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI 3571179).

Ebsworth, M., & Lalumière, M. L. (2012). Viewing time as a measure of bisexual interest. *Archives of Sexual Behavior, 41,* 161–172. doi:10.1007/s10508-012-9923-9

*Fromberger, P., Jordan, K., Steinkrauss, H., von Herder, J., Witzel, J., Stolpmann, G., Kröner-Herwig, B., & Müller, J. L. (2012). Diagnostic accuracy of eye movements in assessing pedophilia. *Journal of Sexual Medicine, 9,* 1868-1882. doi: 10.1111/j.1743-6109.2012.02754.x

Giotakis, O. (2005). A combination of viewing reaction time and incidental learning task in child molesters, rapists, and control males and females. *Sexologies, 14,* 13-20.

*Glasgow, D. V. (2009). Affinity: The development of a self-report assessement of paedophile sexual interest incorporating a viewing time validity measure. In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in sexual offenders* (pp. 59-84). Chichester, UK: Wiley-Blackwell. doi:10.1002/9780470747551.ch3

Gray, S. R., & Plaud, J. J. (2005). A comparison of the Abel Assessment for Sexual Interest and penile plethysmography in an outpatient sample of sexual offenders. *Journal of Sexual Offender Civil Commitment: Science and Law, 1,* 1-10.

Gray, S. R., Abel, G. G., Jordan, A., Garby, T., Wiegel, M., & Harlow, N. (2015). Visual Reaction Time^{TM} as a predictor of sexual offense recidivism. *Sexual Abuse: A Journal of Research and Treatment, 27,* 173-188. doi: 10.1177/1079063213502680

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464-1480. doi:10.1037//0022-3514.74.6.1464

*Gress, C. L. Z. (2005). Viewing time measures and sexual interest: Another piece of the puzzle. *Journal of Sexual Aggression, 11,* 117–125. doi:10.1080/13552600500063666

*Gress, C. L. Z., Anderson, J. O., & Laws, R. D. (2013). Delays in attentional processing when viewing sexual imagery: The development and comparison of two measures. *Legal and Criminological Psychology, 18,* 66-82. doi: 10.1111/j.2044-8333.2011.02032.x

Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (user report #2007-05). Ottawa, ON: Public Safety Canada.

Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73,* 1154-1163. doi: 10.1037/0022-006X.73.6.1154

*Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy, 34,* 389-394. doi:10.1016/0005-7967(95)00070-4

Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment, 4,* 502–511. doi:10.1037/1040-3590.4.4.502

Hasselblad, V. & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117,* 167-178. doi: 10.1037/0033-2909.117.1.167

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560. doi: 10.1136/bmj.327.7414.557.

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples - A meta-analysis. *Criminal Justice and Behavior, 39,* 1148-1171. doi: 10.1177/0093854812443648

Imhoff, R., Schmidt, A. F., Nordsiek, U., Luzar, C., Young, A. W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior*, *39*, 1275–1288. doi:10.1007/s10508-009-9595-2

Imhoff, R., Schmidt, A. F., Weiß, S., Young, A. W., & Banse, R. (2012). Vicarious viewing time: Prolonged response latencies for sexually attractive targets as a function of task- or stimulus-specific processing. *Archives of Sexual Behavior, 41,* 1389–1401. doi:10.1007/s10508-011-9879-1

Israel, E., & Strassberg, D. S. (2009).Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior, 38,* 551–558. doi:10.1007/s10508-007-9246-4

Kalmus, E., & Beech, A. R. (2005).Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10,* 193–218. doi:10.1016/j.avb.2003.12.002

Landis J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics. 33*, 159–174. doi:10.1016/j.chiabu.2008.07.004

*Lanham, D. (2011). *The sensitivity and specificity of phallometric assessment and the Abel Assessment for Sexual Interest among a group of sexual offenders against children and other sexual offenders* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI 3501894)

Larue, D., Schmidt, A. F., Imhoff, R., Eggers, K., Schönbrodt, F., & Banse, R. (2014). Validation of direct and indirect measures of preference for sexualized violence. *Psychological Assessment, 26,* 1173-1183. doi:10.1037/pas0000016

Laws, D. (2009). Penile plethysmography: Strengths, limitations, innovations. In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in sexual offenders* (pp. 7–29). Chichester, UK: Wiley-Blackwell. doi:10.1002/9780470747551.ch1

Lehmann, R. J. B., Goodwill, A. M., Hanson, R. K., & Dahle, K.-P. (2014). Crime scene behaviors indicate risk-relevant propensities of child molesters. *Criminal Justice and Behavior, 41,* 1008-1028. doi: 10.1177/0093854814521807

*Letourneau, E. J. (2002). A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse: A Journal of Research and Treatment, 14,* 207–223. doi:10.1177/107906320201400302

Lippa, R. A. (2012). Effects of sex and sexual orientation on self-reported attraction and viewing times to images of men and women: Testing for category specificity. *Archives of Sexual Behavior, 41,* 149–160. doi:10.1007/s10508-011-9898-y

*Loewinger Cloyd, L. (2007). *Relationship between key variables in penile plethysmograph and viewing time measures of sexual arousal in sex offending adult males* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI 3295152)

*Mackaronis, J. (2014). *Adolescent sexual offending: Assessing sexual interest and exploring treatment trajectories and outcome* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI 3614435)

Mann, R. E., Hanson, K. R., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment, 22,* 191-217. doi:10.1177/1079063210366039

McGrath, R. J., Cumming, G. F., Burchard, B. L., & Ellerby, L. (2010). *Current practices and trends in sexual abuser management: The Safer Society 2009 nationwide survey.* Brandon, Vermont: The Safer Society Press.

*Mokros, A., Gebhard, M., Heinz, V., Marschall, R. W., Nitschke, J., Glasgow, D. V., … Laws, R. D. (2013). Computerized assessment of pedophilic sexual interest through self-report and viewing time: Reliability, validity, and classification accuracy of the Affinity program. *Sexual Abuse: A Journal of Research and Treatment, 25,* 230-258. doi: 10.1177/1079063212454550

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of social cognition – Measurement, theory, and applications* (p. 255-277). New York, NY: Guilford.

Pohl, A., Wolters, A., & Ponseti, J. (2015). Investigating the task dependency of viewing time effects. *Journal of Sex Research.* Advance online publication. doi: 10.1080/00224499.2015.1089429

Prentky, R., & Barbaree, H. (2011). Commetary: Hebephilia – A would-be paraphilia caught in the twilight zone between prepubescence and adulthood. *Journal of the American Academy of Psychiatry and the Law, 39,* 506-510.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association. doi: 10.1037/10304-000

Quinsey, V. L., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology and Sociobiology, 17,* 341–354. doi:10.1016/S0162-3095(96)00060-X

Rice, M. E., Harris, G. T., Lang, C., & Chaplin, T. C. (2012). Adolescents who have sexually offended Is phallometry valid? *Sexual Abuse: A Journal of Research and Treatment, 24,* 133-152. doi: 10.1177/1079063211404249

Rosenzweig, S. (1942).The photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine, 4,* 150–158.

Rönspies, J., Schmidt, A. F., Melnikova, A., Krumova, R., Zolfagari, A. & Banse, R. (2015). Indirect measurement of sexual orientation – Comparison of the Implicit Relational Assessment Procedure, Viewing Time, and Choice Reaction Time Tasks. *Archives of Sexual Behavior, 44,* 1483-1492. doi: 10.1007/s10508-014-0473-1

Ryan, E. P. (2016). Juvenile sex offenders. *Child and Adolescent Psychiatric Clinics of North America, 25,* 81-97. doi:10.1016/j.chc.2015.08.010

*Schmidt, A. F., Bonus, P., & Banse, R. (2010, July). *Indirect measures of sexual interest in child sex offenders: A multimethod approach and its clinical implications*. Paper presented at the International Summer Conference in Forensic Psychiatry, Regensburg, Germany.

Schmidt, A. F., Banse, R., & Imhoff, R. (2015). Indirect measures in forensic contexts. In T. M. Ortner, & F. J. R. van de Vijver (Eds.). *Behavior-based assessment: Going beyond self-*

report in the personality, affective, motivation, and social domains (pp. 173-194). Göttingen: Hogrefe.

*Schmidt, A. F., Gykiere, K., Vanhoeck, K., Mann, R. E., & Banse, R. (2014). Direct and indirect measures of sexual maturity preferences differentiate subtypes of child sexual abusers. *Sexual Abuse: A Journal of Research and Treatment*, *26*, 107–128. doi:10.1177/1079063213480817

Schmidt, A. F., Imhoff, R., & Banse, R. (2014, June). *Viewing time as a measure of sexual interest: A causal explanation of the effect*. Poster presented at the 40th Annual Meeting of the International Academy of Sex Research (IASR), Porto, Portugal.

Schmidt, A. F., Mokros, A., & Banse, R. (2013). Is pedophilic sexual preference continuous? A taxometric analysis based on direct and indirect measures. *Psychological Assessment*, *25*, 1146–1153. doi:10.1037/a0033326

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Zeitschrift für Psychologie / Journal of Psychology, 215*, 90-103. doi: 10.1027/0044-3409.215.2.90

Seto, M. C. (2010). Child pornography use and Internet solicitation in the diagnosis of pedophilia. *Archives of Sexual Behavior, 39,* 591–593. doi: 10.1007/s10508-010-9603-6

Seto, M. C. (2009). Pedophilia. *Annual Review of Clinical Psychology, 5,* 391–407. doi:10.1146=annurev.clinpsy.032408.153618

Seto, M. C., Murphy, W. D., Page, J., & Ennis, L. (2003). Detecting anomalous sexual interests in juvenile sex offenders. *Annals of the New York Academy of Sciences, 989,* 118-130. doi: 10.1111/j.1749-6632.2003.tb07298.x

Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: A Journal of Research and Treatment*, *13*, 15–25. doi:10.1177/107906320101300103

Seto, M. C., Stephens, S., Lalumière, M. L., & Cantor, J. M. (2015). The Revised Screening Scale for Pedophilic Interests (SSPI–2): Development and criterion-related validation. Advance online publication. *Sexual Abuse: A Journal of Research and Treatment.* doi: 1079063215612444.

*Stinson, J. D., & Becker, J. V. (2008). Assessing sexual deviance: A comparison of physiological, historical, and self-report measures. *Journal of Psychiatric Practice, 14,* 379-388. doi: 10.1097/01.pra.0000341892.51124.85

*Weiß, S., Massau, C., Kärgel, C., & Schiffer, B. (2014). *Comparing and combining various latency-based indirect measures of sexual interest*. Poster presented at the 33rd Annual research Conference of the Association for the Treatment of Sexual Abuse (ATSA), San Diego, CA, USA.

*Worling, J. R. (2006). Assessing sexual arousal with adolescent males who have offended sexually: Self-report and unobtrusively measured viewing time. *Sexual Abuse: A Journal of Research and Treatment, 18,* 383–400. doi:10.1177/107906320601800406

Zamansky, H. S. (1956). A technique for measuring homosexual tendencies. *Journal of Personality, 24,* 436–448. doi:10.1111/j.1467-6494.1956.tb01280.x

Table 1.

Overview of Studies Included in the Current Meta-Analysis

	Study	Published (Peer-Review)	Country	SOC Sample Location	SOC Age Category	SOC Treatment Status	*N*				Convergent Validity Data	Scoring Algorithm
							SOC	SOA	NSO	NO		
1	Abel et al. (2004)	Yes (Yes)	USA	Community	Juveniles	Mixed	1170	534	-	-	No	Difference score
2	Babchishin et al. (2014)	Yes (Yes)	Canada	Institution	Adults	Mixed	31	10	20	-	Yes	Difference score
3	Banse et al. (2010)	Yes (Yes)	UK	Institution	Adults	Yes	38	-	37	38[a]	Yes	Aggregate of Difference score and absolute VT for child categories
4	Douroux (2013)	Yes (No)	USA	Community	Adults	Unknown	202	-	-	-	Yes	-
5	Fromberger et al. (2012)	Yes (Yes)	Germany	Institution	Adults	Yes	22	-	-	60[b]	No	Difference score
6	Glasgow (2009)	Yes (Yes)	UK	Institution	Adults	Yes	31	-	-	31	No	Difference score
7	Gress (2005)	Yes (Yes)	Canada	Community	Adults	Yes	19	7	-	-	No	Difference score
8	Gress et al. (2013)	Yes (Yes)	Canada	Community	Adults	Yes	22	-	40[c]	59	No	Absolute VT for child categories
9	Harris et al. (1996)	Yes (Yes)	Canada	Combined	Adults	Unknown	26	-	-	25	Yes	Difference score
10	Lanham (2011)	Yes (No)	USA	Community	Adults	Yes	45	53	-	-	Yes	Absolute VT for child categories

	Study	Published (Peer-Review)	Country	SOC Sample Location	SOC Age Category	SOC Treatment Status	*N*				Convergent Validity Data	Scoring Algorithm
							SOC	SOA	NSO	NO		
11	Letourneau (2002)	Yes (Yes)	USA	Institution	Adults	No	18	35[d]	-	-	Yes	Absolute VT for child categories
12	Loewinger Cloyd (2007)	Yes (No)	USA	Community	Adults	Yes	96		-	-	Yes	-
13	Mackaronis (2014)	Yes (No)	USA	Community	Juveniles	Yes	16	-	-	-	Yes	-
14	Mokros et al. (2013)	Yes (Yes)	Germany	Institution	Adults	Yes	42	-	27	95	Yes	Difference score[f]
15	Schmidt et al. (2010)	No (No)	Germany	Institution	Adults	Yes	45	-	28	-	Yes	Difference score
16	Schmidt et al. (2014)[e]	Yes (Yes)	Belgium	Community	Adults	Yes	54				Yes	Difference score
17	Stinson & Becker (2008)	Yes (Yes)	USA	Community	Adults	Yes	60	-	-	-	Yes	Absolute VT for child categories
18	Weiß et al. (2014)	No (No)	Germany	Unknown	Adults	Unknown	29	-	-	30	No	Difference score
19	Worling et al. (2006)	Yes (Yes)	Canada/ USA	Community	Juveniles	Yes	52	26	-	-	No	Difference score

Note. SOC = sexual offenders against children; SOA = sexual offenders against adults; NSO = non-sexual offenders, NO = non-offenders (community or students); [a] Mixed sample of community and student males; [b] Includes 8 non-pedophilic forensic controls; [c] Adolescent sample (M = 16.5 years of age); [d] Comparison of SOC with girl victims to SOC without girl victims (including SOC with boy victims only)/ comparison of SOC with boy victims to SOC without boy victims (including SOC with girls victims only); [e] Comparison groups were non-contact SOC (users of child exploitative sexual material, n =18). As such, the study was only used for convergent validity meta-analysis; [f] Difference scores based on absolute VTs were used instead of residual scores reported in the paper (raw data obtained from the first author of the original study).

Table 2.

Meta-analysis of Viewing Time Measures of Sexual Interest in Children in Sexual Offenders Against Children vs. Control Groups

Comparison	**Fixed-Effect**		**Random-Effects**					
	d	**95% *CI***	***d***	**95% *CI***	I^2	***Q***	***N* (*k*)**	**Studies**
Overall aggregated effect	0.596	[0.514, 0.679]	0.773	[0.577, 0.969]	66.2%	38.50***	2,705 (14)	1,2,3,5,6,7,8,9,10, 11,14,15,18,19
SOC vs. NO	0.845	[0.659,1.030]	0.920	[0.586, 1.255]	68.2%	18.90**	548 (7)	3,5,6,8,9,14,18
SOC vs. NSO	0.574	[0.376, 0.771]	0.578	[0.370, 0.785]	9.0%	5.50	428 (6)	2,3,8,10,14,15
SOC vs. SOA	0.545	[0.447, 0.643]	0.765	[0.404, 1.126]	64.3%	11.21*	1,912 (5)	1,2,7,11,19
SOC vs. SOA[a] (outlier removed)	0.524	[0.425, 0.624]	0.594	[0.329, 0.860]	35.5%	4.65	1,859 (4)	1,2,7,19

Note. SOC = sexual offenders against children; SOA = sexual offenders against adults; NSO = non-sexual offenders, NO = non-offenders (community or students); *k* = number of samples. [a] Excluding one outlying study (Letourneau, 2002) where SOC were included in the comparison group. Despite study #1 (Abel et al., 2004) having a sample size much larger than the other included studies, reducing its study weight produced remarkably similar results (likely due to it being in the middle of the distribution). As such, non-transformed data are presented.

* *p* < .05, ** *p* < .01, *** *p* < .001

Table 3.

Moderator Analyses for Control Group Studies of Viewing Time Measures of Sexual Interest in Children

	Fixed-Effect		**Random-Effects**					
Moderator	***d***	**95% *CI***	***d***	**95% *CI***	***I*²**	***Q***	***N* (*k*)**	**Studies**
Sample Age Category								
Juveniles	0.505	[0.403, 0.606]	0.505	[0.403, 0.606]	0%	0.22	1,782 (2)	1,19
Adults	0.779	[0.636, 0.922]	0.857	[0.617, 1.097]	61.9%	28.91**	923 (12)	2,3,5,6,7,8,9,10,11,14,15,18
						Q-between* 9.37*		
Scoring Method[a]								
Absolute Viewing Time Score	0.630	[0.410, 0.850]	0.701	[0.333, 1.069]	61.5%	7.79	396 (4)	8,10,11,14
Difference Score	1.034	[0.819, 1.250]	1.047	[0.758, 1.250]	42.2%	10.38	414 (7)	2,5,6,7,9,15,18
						Q-between* 24.81**		
Stimuli Type[b]								
Nude Pictures	1.093	[0.778, 1.408]	1.093	[0.778, 1.408]	0%	0.74	194 (3)	2,5,9
Nude and Clothed	0.809	[0.349, 1.268]	0.809	[0.349, 1.268]	0%	0.01	107 (2)	7,8
Clothed Pictures	0.682	[0.510, 0.854]	0.789	[0.444, 1.134]	74.0%	23.10***	622 (7)	3,6,10,11,14,15,18
						Q-between 5.06‡		
Type of Pictures								
Computed-Generated Photos	0.908	[0.682, 1.135]	0.914	[0.658,1.170]	19.5%	6.21	382 (6)	2,5,7,8,15,18
Real Photos	0.775	[0.566,0.985]	0.922	[0.461,1.384]	77.6%	17.88**	428 (5)	6,9,10,11,14
						Q-between	0.72	
Include only prepubescent stimuli								
Yes	1.029	[0.636, 1.422]	1.051	[0.534,1.569]	40.7%	1.69	134 (2)	8,11
No	0.741	[0.587, 0.894]	0.821	[0.552, 1.089]	64.6%	25.43**	789 (10)	2,3,5,6,7,9,10,14,15,18
						Q-between 1.78		
Publication Bias								
Published	0.782	[0.626, 0.938]	0.873	[0.600, 1.147]	64.6%	25.46**	791 (10)	2,3,5,6,7,8,9,10,11,14
Unpublished	0.763	[0.402, 1.125]	0.798	[0.122, 1.475]	70.9%	3.44	132 (2)	15,18
						Q-between 0.01		
Peer Reviewed	0.787	[0.609, 0.964]	0.849	[0.591, 1.106]	47.9%	13.43**	631 (8)	2,3,5,7,8,9,11,14
Non-Peer Reviewed	0.765	[0.522, 1.007]	0.869	[0.308, 1.431]	80.6%	15.46	292 (4)	6,10,15,18
						Q-between 0.02		

Criteria for SOC group								
Index is SOC	0.878	[0.665, 1.091]	0.984	[0.562, 1.406]	71.0%	17.22**	439 (6)	2,6,7,11,14,15
Any history of SOC	0.697	[0.503, 0.890]	0.745	[0.464, 1.026]	50.8%	10.16	484 (6)	3,5,8,9,10,18
						Q-between 1.53		

Note. SOC = sexual offending against children. The two juvenile samples (studies 1, 19) were removed from all subsequent moderator analyses as age was a statistically significant moderator of group discrimination. A Q-between with asterisks represents a statistically significant moderator. [a] One study which was an aggregate effect size of difference scores and absolute viewing times (study 3) was removed from this analysis. [b] *p* = .079 for Stimuli Type.
ǂ < .10, ** *p* < .01, *** *p* <.001

Table 4.

Meta-Analyses of Convergent Validity with External Criteria of Sexual Interest in Children

Measure	**Fixed-Effect**		**Random-Effects**					
	r	**95% *CI***	***r***	**95% *CI***	***I*²**	***Q***	***N* (*k*)**	**Studies**
IAT	.181	[.062, .295]	.181	[.062, .295]	0%	1.70	323 (4)	2,3,15,16
PPG	.248	[.144, .346]	.160	[-.066, .371]	73.6%	18.95**	347 (6)	2,10,11,12,13,17
Self-report	.375	[.285, .459]	.375	[.285, .459]	1.9%	6.12	397 (7)	2,3,6,9,15,16,17
SSPI	.212	[.122, .299]	.183	[.001, .354]	68.4%	18.98**	429 (7)	2,3,4,13,14,15,16

Note. Meta-analyses conducted on Fisher Z; retransformed data are presented for ease of interpretation. SSPI = Screening Scale for Pedophilic Interests (Seto & Lalumière, 2001); PPG = penile plethysmography; IAT = Implicit Association Test.

** $p < .01$

Figure 1. Forest plot of studies included in the comparison group meta-analysis and the weighted fixed-effect average. Displayed are effect sizes (Cohen's *d*) with their 95% confidence intervals. The effects crossing the vertical line did not reach statistical significance at $p < .05$.

Figure 1

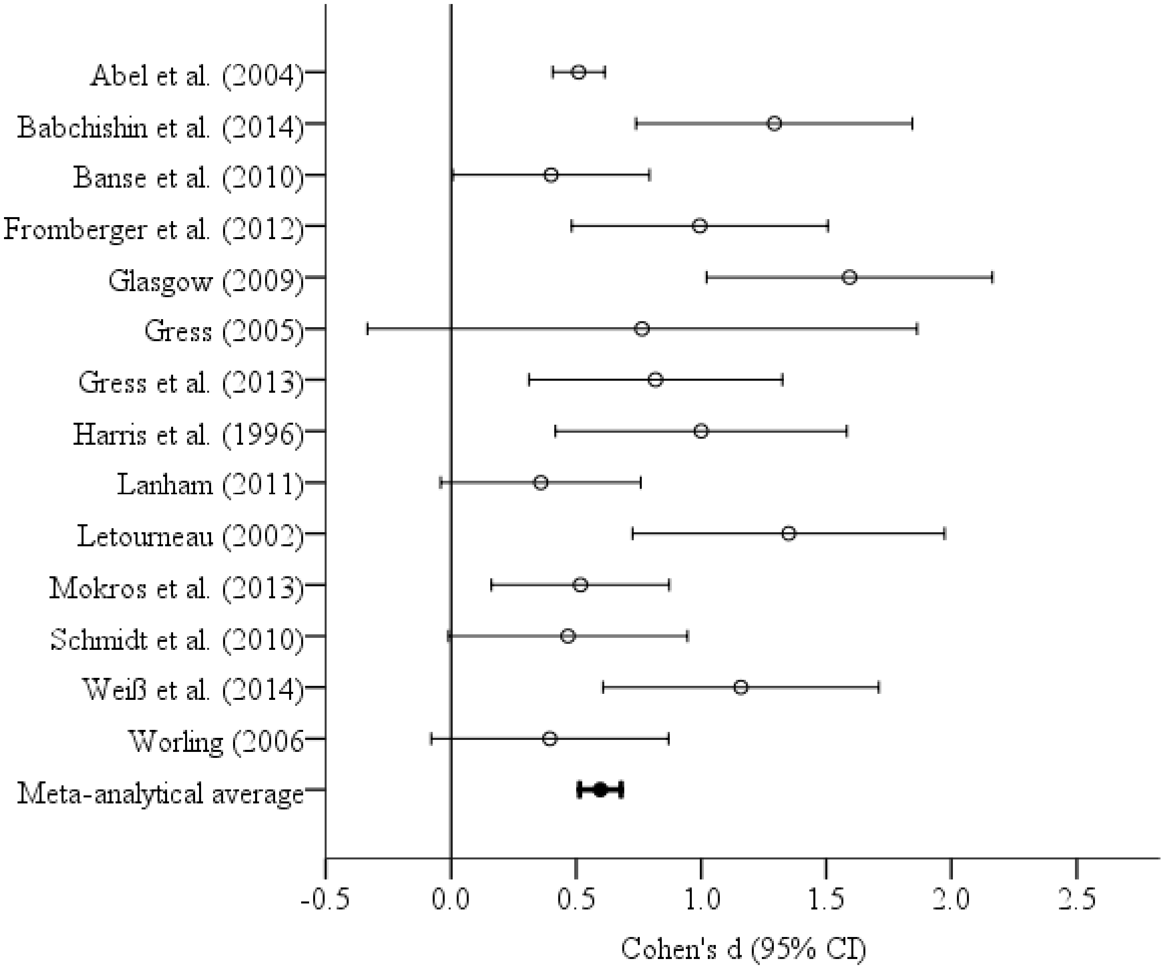
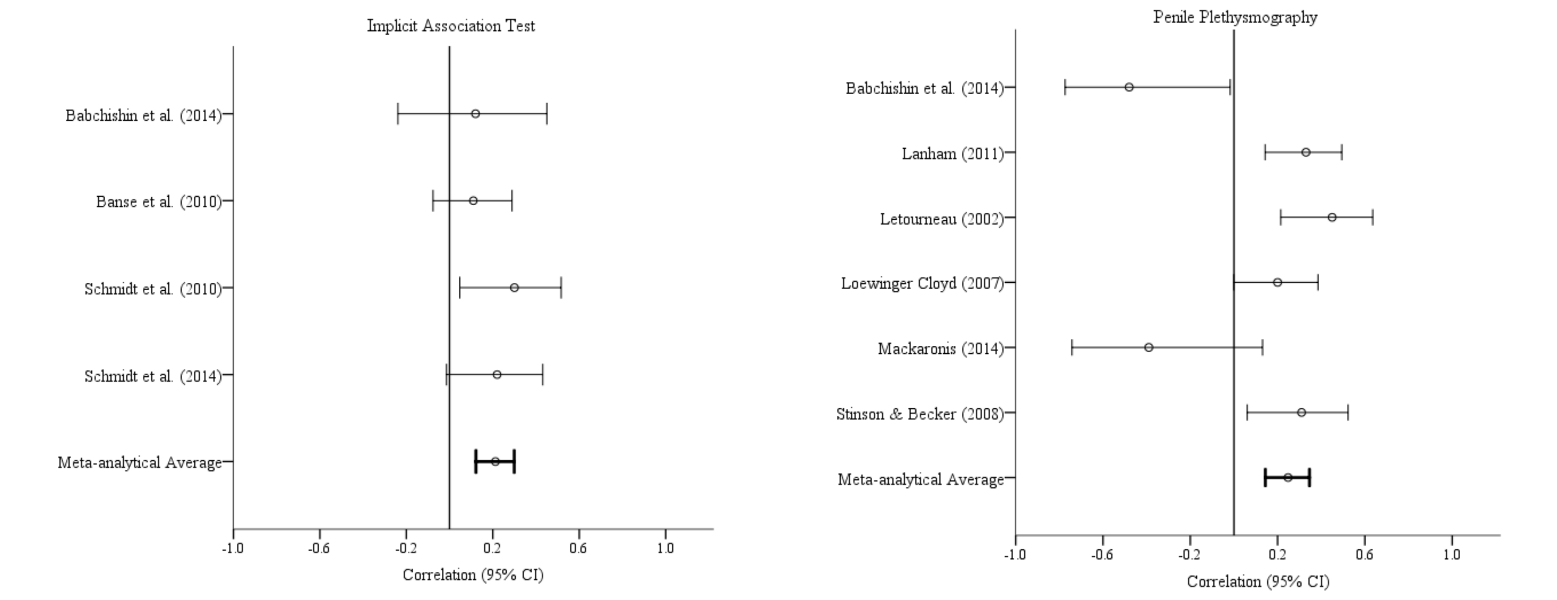
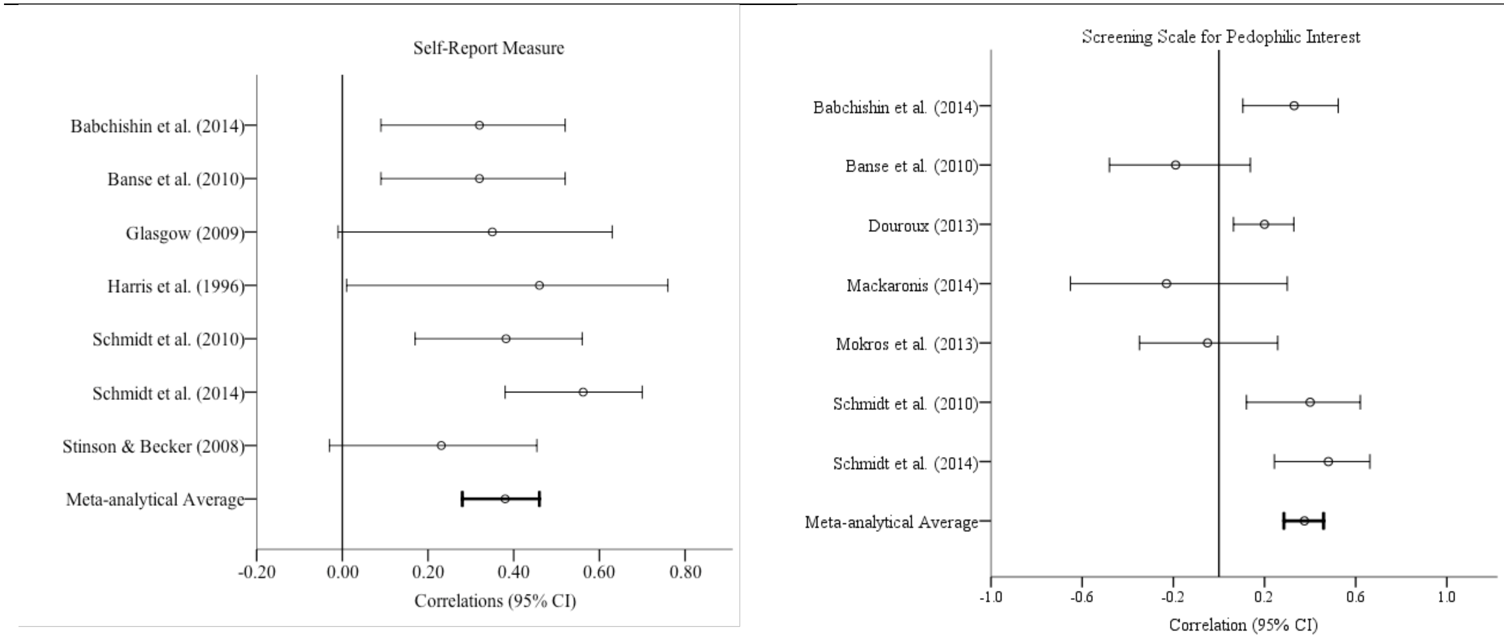

Abel et al. (2004)
Babchishin et al. (2014)
Banse et al. (2010)
Fromberger et al. (2012)
Glasgow (2009)
Gress (2005)
Gress et al. (2013)
Harris et al. (1996)
Lanham (2011)
Letourneau (2002)
Mokros et al. (2013)
Schmidt et al. (2010)
Weiß et al. (2014)
Worling (2006
Meta-analytical average
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
Cohen's d (95% CI)

Figure 2. Forest plots of studies included in the convergent validity meta-analysis with criterion measures of sexual interest in children and the weighted fixed-effect average. Displayed are effect sizes (*r*) with their 95% confidence intervals. Correlations were transformed from Fisher's Z. The effects crossing the vertical line did not reach statistical significance at *p* < .05.

Figure 2


Implicit Association Test
Babchishin et al. (2014)
Banse et al. (2010)
Schmidt et al. (2010)
Schmidt et al. (2014)
Meta-analytical Average
-1.0
-0.6
-0.2
0.2
0.6
1.0
Correlation (95% CI)
Penile Plethysmography
Babchishin et al. (2014)
Lanham (2011)
Letourneau (2002)
Loewinger Cloyd (2007)
Mackaronis (2014)
Stinson & Becker (2008)
Meta-analytical Average
-1.0
-0.6
-0.2
0.2
0.6
1.0
Correlation (95% CI)


Self-Report Measure
Babchishin et al. (2014)
Banse et al. (2010)
Glasgow (2009)
Harris et al. (1996)
Schmidt et al. (2010)
Schmidt et al. (2014)
Stinson & Becker (2008)
Meta-analytical Average
-0.20
0.00
0.20
0.40
0.60
0.80
Correlations (95% CI)
Screening Scale for Pedophilic Interest
Babchishin et al. (2014)
Banse et al. (2010)
Douroux (2013)
Mackaronis (2014)
Mokros et al. (2013)
Schmidt et al. (2010)
Schmidt et al. (2014)
Meta-analytical Average
-1.0
-0.6
-0.2
0.2
0.6
1.0
Correlation (95% CI)