Article

Sexual Abuse: A Journal of Research and Treatment
25(3) 230–258

DOI: 10.1177/1079063212454550
sax.sagepub.com
SAGE

Computerized Assessment of Pedophilic Sexual Interest Through Self-Report and Viewing Time: Reliability, Validity, and Classification Accuracy of the Affinity Program

Andreas Mokros[1,2], Michael Gebhard[2], Volker Heinz[2], Roland W. Marschall[3], Joachim Nitschke[3], David V. Glasgow[4], Carmen L. Z. Gress[4], and D. Richard Laws[4]

Abstract

Affinity is a computerized assessment tool that combines viewing time and self-report measures of sexual interest. The present study was designed to assess the diagnostic properties of Affinity with respect to sexual interest in prepubescent children. Reliability of both self-report and viewing time components was estimated to be high. The group profile of a sample of pedophilic adult male child molesters (n = 42, all of whom admitted their offenses) differed from the group profiles of male community controls (n = 95) and male nonsexual offenders (n = 27), respectively. More specifically, both ratings and viewing times for images showing small children or prejuvenile children were significantly higher within the child molester sample than in either of the other two groups, attesting to the validity of the measures. Overall classification accuracy, however, was mediocre: A multivariate classification routine

[1]University Hospital of Psychiatry Zurich, Zurich, Switzerland
[2]University of Regensburg, Regensburg, Germany
[3]Ansbach District Hospital, Ansbach, Germany
[4]Pacific Psychological Assessment Corporation, Victoria, British Columbia, Canada

Corresponding Author:
Andreas Mokros, University Hospital of Psychiatry Zurich, Center for Forensic Psychiatry, Lenggstrasse 31, P. O. Box 1931, Zurich, CH-8032, Switzerland.
Email: andreas.mokros@puk.zh.ch

yielded 50% sensitivity for child molester status at the cost of 13% false positives. The implications for forensic use of Affinity are discussed.

Keywords

sexual interest, viewing time, self-report, adult, pedophilia

Implicit Measurement of Pedophilic Sexual Interest

Recently, cognitive methods that rely on shifts in reaction time have come under investigation as potential diagnostic tools for deviant sexual interest (Gress & Laws, 2009; Snowden, Craig, & Gray, 2011; Thornton & Laws, 2009). There is a considerable body of evidence suggesting that pedophilic men have a processing bias for child-related stimuli. Pedophilic participants need longer reaction times or make more errors in response tasks that involve child stimuli. Among the experimental paradigms that indicated such a processing bias for child stimuli among pedophiles were *choice reaction time* (CRT; Giotakos, 2005, 2006; Gress, 2007; Mokros, Dombert, Osterheider, Zappalà, & Santtila, 2010; Poeppl et al., 2011), the Implicit Association Test (IAT; Banse, Schmidt, & Clarbour, 2010; Gray, Brown, MacCulloch, Smith, & Snowden, 2005; Mihailides, Devilly, & Ward, 2004; Nunes, Firestone, & Baldwin, 2007; Steffens, Yundina, & Panning, 2008), modified versions of the Stroop task (Ó Ciardha & Gormley, 2012; Price & Hanson, 2007; Smith & Waterman, 2004), *rapid serial visual presentation* (RSVP; Beech et al., 2008) and *viewing time* (e.g., Abel, Lawry, Karlstrom, Osborn, & Gillespie, 1994; Banse et al., 2010; Harris, Rice, Quinsey, & Chaplin, 1996).

According to the information processing model of Janssen, Everaerd, Spiering, and Janssen (2000) these results would imply that pedophilic men were particularly prone to direct their attention toward child-related stimuli in an automatic, unconscious way. This accords well with the view of automatic processes of mate selection leading to longer reactions times (Maner, Gailliot, Rouby, & Miller, 2007). Possibly, such mate selection processes follow a sequential pattern (Imhoff et al., 2010), involving criteria such as stimulus age, gender, and physical attractiveness (not necessarily in this order). Alternatively, sexually interesting stimuli may lead to a conscious allocation of cognitive resources that would impede with the performance in other simultaneous tasks (see Kahneman, 1973). Both perceptual attention (i.e., selection of relevant targets among distracters) and response selection (i.e., deciding between actions) are impaired the more demanding the task becomes, for example, by increasing the amount of alternatives (Pashler, 1998). This would seem a plausible explanation for longer response latencies in dual tasks that involve sexual stimuli (such as in the CRT, modified Stroop or RSVP paradigms). Yet another interpretation is the notion of increased cognitive monitoring with regard to sexual stimuli—a phenomenon that Geer and Melton (1997) termed sexual content–induced delay (SCID). Consequently, it is either an automatic link or a conscious preponderance (attention bias or increased response monitoring) that leads to longer response latencies in the presence of sexual stimuli.

Imhoff et al. (2010) presented a series of experiments in which viewing time effects were shown to depend on stimulus attractiveness (based on varying age and gender of persons shown) even under restricted task conditions, such as time pressure or limited stimulus displays. Jiang, Costello, Fang, Huang, and He (2005) observed gender- and sexual orientation–specific differences in the spatial allocation of attention: Hetero- and homosexual male and female participants performed worse in a visual discrimination task in the presence of sexually relevant images, even though these images went unnoticed on a conscious level due to interocular suppression. Jiang et al. interpreted this finding in terms of selective attention that would be affected by the sexual relevance of a visual stimulus in an automatic way. Interestingly, the effect vanished if the stimuli were presented supraliminally for 800 milliseconds so that the participants became aware of them.

Based on the findings by Imhoff et al. (2010) and by Jiang et al. (2005) it seems tenable that the assessment of sexual interest based on differences in response times toward pertinent visual stimuli qualifies as an implicit measure according to the definition given by De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009).[1] According to the interpretation of the viewing time assessment of pedophilic sexual interest as an implicit measure, pedophilic individuals would share a distinctive automatic processing bias for child-related stimuli that would lead to prolonged response times in the presence of such stimuli. The notion of an automatic bias would follow from the conceptualization of the sexual arousal reaction by Singer (1984) with an aesthetic response of attention allocation as its first stage. This view remains tentative, however, because the causal mechanisms are not yet fully understood. Furthermore, variations in the measurement outcome may also be due to other confounding factors, such as disgust or curiosity.

Viewing Time Assessment

The idea that the covert registration of the time that a person looks at assorted pictures of varying sexual content could serve as a proxy for sexual interest dates back to earlier research by Rosenzweig (1942) and Zamansky (1956). The status of viewing time assessment as an implicit measurement procedure makes it relevant for forensic clinical assessment of deviant sexual interest because viewing time assessment may be less susceptible to deliberate dissimulation than self-report questionnaires, for instance. According to comparative studies of viewing time assessments with the erectile arousal response (as measured through penile plethysmography, PPG) in child molesters there were similar levels of reliability and criterion validity for both methods in some studies (Abel, Huffman, Warberg, & Holland, 1998; Abel, Jordan, Hand, Holland, & Phipps, 2001; Gray & Plaud, 2005), but mixed results in other studies (Letourneau, 2002; Stinson & Becker, 2008). Harris et al. (1996) noted a large effect size ($d = 1.0$) for viewing time assessments of child sexual abusers that fell below the effect size of PPG ($d = 2.1$), however.

Currently, there are several computer programs commercially available that partly rely on viewing time in order to assess deviant sexual interest, including the Abel Assessment for Sexual Interest—2™ (AASI-2; Abel Screening, Inc., n.d.), Visual

Sexual Preference Assessment (VSAP; Limestone Technologies, Inc., n.d.), and Affinity (Pacific Psychological Assessment Corporation, Inc., 2007). The Affinity software is described in detail by Glasgow (2009). The original publication by Glasgow, Osborne, and Croxen (2003) on Affinity contains five case examples for purposes of illustration. Apart from unobtrusively recorded viewing time, Affinity also incorporates a self-report feature with respect to sexual interest: For every picture shown to the participant, he or she has to indicate the extent to which he or she regards the image as sexually attractive. A first evaluation of a previous version of the Affinity program (based on a total of 56 images; 7 images per category of age and gender) with 78 admitting child and juvenile sexual offenders (at ages 12 to 18) by Worling (2006) yielded average estimates of internal consistency of Cronbach's $\alpha = .76$ (range: .62-.82) for the viewing time feature and $\alpha = .95$ (range: .87-.97) for the image rating task. The correlations between image ratings and viewing times across the four age categories of male and female images were moderate (with a mean correlation coefficient of $r = .31$). All pairs of viewing time and image ratings per picture category (as defined through age and gender) correlated positively to a significant degree, with the exception of images of adult women ($r = -.26$). Based on the ratio between responses toward pictures of (pre)pubescent versus juvenile or adult individuals, both the viewing time feature and the self-report task yielded medium effect sizes when trying to distinguish between subgroups within the sample (any child victim, multiple child victims, male victims).

Apart from the introductory article by Glasgow et al. (2003) and the validation study by Worling (2006), published empirical studies on Affinity are lacking. There are a couple of unpublished dissertations dealing with Affinity, however (Brown, 2005; Croxen, 2003; Loewinger Cloyd, 2007; Worsham, 2010). Croxen (2003) compared the viewing time profiles of 31 male child sexual abusers with those of 31 male nonoffender controls. Within the child sexual abuser group, nearly half of the participants (14 out of 31 individuals) in Croxen's study denied any sexual attraction toward children in a self-report questionnaire, whereas the remainder of this group admitted to at least some sexual attraction toward children. None of the control sample indicated any sexual interest in children in the questionnaire. Using the same version of the Affinity program as Worling (2006), Croxen (2003) noted internal consistency estimates (Cronbach's α coefficients) ranging from .76 to .93. On a differential index of pedophilic interest (subtracting the maximum viewing times for the adult categories from the maximum viewing time for small children or juveniles), the mean of the child sexual abusers was 1.59 standard deviation units above the mean for the controls—a large and significant effect.[2] In Croxen's study, choosing a cutoff that classified all but one child sexual abuser as pedophilic (96.8% sensitivity) led to 24 out of 31 controls allocated correctly to the nonpedophilic group (77.4% specificity)—an overall hit rate of 87.1% that was clearly above the 50% margin of hits expected by chance.[3]

All in all, prior research seems to show that viewing time and self-report assessment of deviant sexual interest may be useful for the diagnosis and treatment of child sexual abusers. Theoretically, it is plausible to assume that pedophilic individuals will

display a biased appraisal of visual child stimuli. Presently, results on the reliability, the criterion validity, and the fakability of viewing-time methods are comparatively scarce. There are indications of concurrent validity with physiological assessments of deviant sexual arousal (Abel et al., 1998; Harris et al., 1996; Laws, Hanson, Osborn, & Greenbaum, 2000; Letourneau, 2002) and with sexual offense history data, however (Abel et al., 2004; Banse et al., 2010; Worling, 2006; but also see Stinson & Becker, 2008).

The aims of the current study were threefold: First, the objective was to estimate the internal consistency and split-half reliability of the picture categories within Affinity 2.5 for both the viewing time feature and for the image rating task. Second, we aimed to assess the validity of Affinity 2.5 in terms of criterion validity (distinguishing child sexual abusers from nonsexual offender and community controls) as well as concurrent and discriminant validity. We hypothesized that

Hypothesis 1: There would be a two-way interaction for image ratings and viewing times with participants' group membership (child sexual abusers vs. nonsexual offenders vs. community controls) in separate ANOVAs (analyses of variance), with child sexual abusers showing longer response latencies and more positive ratings toward the (pre)pubescent picture categories of small children and prejuveniles (criterion validity).

Hypothesis 2: within one image category (described by age and gender) average viewing times would be correlated positively with image ratings; furthermore, a biographical index of sexual offending against children would correlate positively with differential indices of sexual preference for children derived from viewing times and image ratings for pictures of small children and of adults, respectively (concurrent validity).

Hypothesis 3: The assessment of sexual interest should not be confounded with either intellectual ability or with a socially desirable response style; hence, viewing time and image ratings would be linearly independent of verbal intelligence and the proneness toward socially desirable responding (discriminant validity).

Child sexual abusers, in particular those who were likely pedophilic, had lower IQ scores than nonsexual offenders and nonoffenders according to a meta-analysis (Cantor, Blanchard, Robichaud, & Christensen, 2005). Still, a measure of pedophilic sexual interest should not correlate substantially with intelligence for picture categories showing individuals of particular age groups. Otherwise, the measure would likely pick up cognitive processing speed or test knowledge rather than the intended criterion.

Third, we attempted to compare classification procedures for identifying likely pedophilic sexual interest based on Affinity 2.5. On the basis of the observation that composite indices of deviant sexual interest within Affinity showed higher diagnostic utility (Worling, 2006), we hypothesized that a multivariate index would be more accurate than univariate differential indices.

Method

Participants

A total of 164 men took part in the present study, 42 of whom had been convicted of hands-on sexual offenses against children (child sexual abusers). The participants within the child sexual abuser group acknowledged their offenses. These offenders had sexually abused on average four victims each (SD = 3.83, range: 1-20, median = 3). Primarily, the child sexual abusers were extrafamilial offenders (6/7 of the subsample or 36 individuals). Counting only the youngest child in the case of multiple victims, the mean age of the victims of the child sexual abusers had been 8.65 years (SD = 3.03, range: 2-16, median = 8 years).[4] The victims of the child sexual abusers had been predominantly male (about 3/5 of the subsample or 25 individuals), with one offender victimizing both boys and girls. All members of the child sexual abuser subgroup were inpatients at one of five medium-to-high security forensic-psychiatric hospitals (located at Ansbach, Erlangen, Mainkofen, Regensburg, and Straubing, Germany) under mandatory treatment order and diagnosed with pedophilia according to the ICD-10 criteria of the World Health Organization (2004). Another 27 participants who did not have any previous conviction for a sexual offense and no diagnosis of any paraphilia were also sampled from the same forensic-psychiatric hospitals (nonsexual offender group). Exclusion criteria for all patients were psychosis, treatment with neuroleptic medication, attention deficit disorder, or a noncorrected vision impairment. Patients were sampled at various stages of treatment, with detainment orders generally lasting several years. The remaining 95 individuals were sampled from among patients and visitors at a local general (nonpsychiatric) university hospital (community controls). Participants gave their informed consent prior to taking part in the study. The research was approved by the Ethics Committee of the School of Medicine at the University of Regensburg, Germany.

Design and Procedure

Apparatus. Participants were tested individually using Affinity 2.5 (Pacific Psychological Assessment Corporation, Inc., 2007; see Glasgow, 2009, for details). The program was run on laptop computers with 14-inch thin-film-transistor displays. Affinity uses a total of eighty photographs (40 images of women, 40 images of men), depicted fully clothed in frontal poses, within natural surroundings. None of the pictures are pornographic or sexually suggestive. The picture set comprises 10 photographs from each of 8 categories of age and gender: small children, prejuveniles, juveniles, and adults. The age categories roughly correspond to children 5 years of age and younger (small children), children between the ages of 6 and 10 years (prejuveniles), pubescents, and adolescents between the ages of 11 and 15 years (juveniles), and adults 18 years and above. An initial set of 56 of these images had been chosen based on perfect agreement among five mental health professionals who had allocated the pictures to age groups independently (Glasgow, 2009). The remaining pictures (three per category of age and gender) were added later on to match the present ones.

In the Affinity program, the pictures are presented in a random order. Upon showing each picture, the participant is asked to indicate whether he or she regards the person depicted as sexually attractive (based on a visual analogue scale ranging from *unattractive* to *attractive*, each spanning seven gradients and including a neutral center). The task of judging each picture in terms of attractiveness will henceforth be referred to as *image rating*. The participant uses the computer mouse to record his or her ratings. Unbeknownst to the participant, the *viewing time* between the stimulus onset and the response is also registered (in 100th of a second). The focus of the current study is on these viewing time and image rating features of Affinity.

Psychometric Measures

Additional data were collected for subgroups of participants. For all participants from the child sexual abuser and from the nonsexual offender groups, the judicial verdicts and crime registry sheets were content-analyzed with respect to previous convictions. For subsamples of these two groups (21 child sexual abusers and 15 nonsexual offenders) as well as for the entire sample of community controls further data (on verbal intelligence, personality, and sexual orientation) were collected. In these cases, participants took a brief intelligence test, filled in a self-report personality questionnaire, and answered two items on sexual orientation in writing after the Affinity testing had been conducted.

Screening Scale for Pedophilic Interests. The Screening Scale for Pedophilic Interests (SSPI; Seto & Lalumière, 2001) is a brief actuarial instrument that comprises four items (male victim, multiple victims, victim younger than 12 years, unrelated victim). Each of these items is scored with 1 if present (except the item *male victim* that affords a score of 2) based on an individual's sexual offense history, yielding a maximum total score of 5. The SSPI total score has been shown to correlate with pedophilic sexual arousal (Seto & Lalumière, 2001), pedophilic sexual interest (Banse et al., 2010; Mokros, Dombert et al., 2010), and sexual offense recidivism (Seto, Harris, Rice, & Barbaree, 2004). Due to its focus on offense and victim characteristics, the SSPI can only be applied to individuals with at least one previous conviction for a sexual offense against a minor (Seto & Lalumière, 2001). Although reliability estimates for the SSPI have not been published in the literature yet, the correlation of the SSPI with a questionnaire on sexual interest in boys ($r = .58$) reported by Banse et al. (2010) allows estimating the lower bound for the reliability of the SSPI as $r_{tt} \geq .38$ (see Lord & Novick, 1968).[5]

Multiple-Choice Vocabulary Test. The Multiple Choice Vocabulary Test (MWT-B) by Lehrl (1999; see also Lehrl, Triebig, & Fischer, 1995) is a German-language screening test of verbal intelligence. The test consists of 37 items. The task is to identify proper nouns among distracters. According to the test manual (Lehrl, 1999), the test–retest reliability coefficient is given as $r_{tt} = .95$ (after 30 min) and as $r_{tt} = .87$ (after 14 months). The median correlation coefficient of the MWT-B with 32 other global tests of intelligence is reported as $r = .72$. The MWT-B predominantly

differentiates in the lower to medium IQ range, and the error of measurement is larger in the upper range.

Self-report personality inventory. The German version (Ruch, 1999) of the Eysenck Personality Questionnaire—Revised (EPQ-R; Eysenck, Eysenck, & Barrett, 1985) is a personality inventory that consists of 102 items with a dichotomous (yes/no) answer format. The items represent four scales: Extraversion (E, 23 items), Neuroticism (N, 25 items), Psychoticism (P, 32 items), and a Lie scale of socially desirable responding (L, 22 items). In a sample of men ($N = 82$) who filled in the EPQ-R twice within up to 6 weeks (Ruch, 1999), the test–retest reliability coefficients of the scales were estimated at r_{tt} = .88 (E), .86 (N), .85 (P), and .81 (L).

Sexual orientation. Sexual orientation was ascertained through German translations of two questions on sexual identity from the Sell Assessment of Sexual Orientation (SASO; Sell, 1996): "I consider myself homosexual . . ." and "I consider myself heterosexual . . .", with ratings ranging from *not at all* to *extremely* on a 7-point scale. According to Sell (1996) test–retest reliability for these two items in an online survey was r_{tt} = .93 and .94 in a sample of 152 men after an interval of about 2 weeks.

Data Analysis

To control for possible outliers in viewing time and image rating data, individual responses per picture category were indexed by the respective median value. In keeping with the method for calculating a pedophilic index in PPG research (Harris, Rice, Quinsey, Chaplin, & Earls, 1992; see also Blanchard, Klassen, Dickey, Kuban, & Blak, 2001) we used the maximum median values in viewing times per age category, irrespective of the gender of the persons displayed on the images. In this way, the sexual orientation of the participants in terms of gender was removed from the analyses. The low count of homosexual participants (especially in the nonsexual offender group and among the community controls) precluded the formal analysis of a Stimulus Gender × Participants' Sexual Orientation interaction.

For purposes of classification, differential indices of pedophilic interest were calculated by subtracting the higher median for an adult category (male or female) from the higher median for a small children category (boy or girl), for both viewing times and image ratings. The diagnostic properties of these two univariate indices of preference were compared with a multivariate normal Bayesian classification algorithm (Tatsuoka, 1971; Wald, 1944; see Mokros, Stadtland, Osterheider, & Nedopil, 2010, for details).

Effect sizes were described with absolute mean differences in standard deviation units (i.e., Cohen's *d* coefficient; Cohen, 1992) or in terms of the coefficient denoting the area under a receiver operating characteristic curve (ROC), the AUC (area under the curve) coefficient. The area underneath an ROC curve ($0 \leq AUC \leq 1$) denotes how well the test in question differentiates between cases and noncases. AUC values above .5 indicate that across the whole range of possible cutoff values the test yields a higher true-positive than false-positive rate. In order to test the difference between two

Table 1. Group Means (*SDs*) for Age, Verbal IQ, and the Lie Scale.

	Child Sexual Abusers	Nonsexual Offenders	Community Controls
Age	42.69	32.44	35.60
	(10.44)	(10.10)	(13.04)
	n = 42	*n* = 27	*n* = 95
Verbal IQ	101.62	97.75	109.85
	(12.88)	(7.07)	(14.27)
	n = 21	*n* = 16	*n* = 95
Lie scale	8.29	4.31	7.02
	(5.23)	(4.44)	(4.95)
	n = 21	*n* = 16	*n* = 90

Note: Age was participants' age at the time of the assessment. Verbal IQ was measured with the Multiple Choice Vocabulary Test (MWT-B) by Lehrl (1999; see also Lehrl, Triebig, & Fischer, 1995). The Lie scale refers to the control scale for socially desirable response style from the Eysenck Personality Questionnaire—Revised (EPQ-R; Eysenck, Eysenck, & Barrett, 1985), German version by Ruch (1999).

ROC-AUCs derived from the same data for statistical significance, we used the method developed by DeLong, DeLong, and Clarke-Pearson (1988).

Results

Descriptive Analyses

Participants were between 18 and 75 years of age (M = 36.90, SD = 12.44). Their average verbal IQ (based on a subsample of 129 cases) ranged from 77 to 145 points (M = 107.01, SD = 14.07). For 127 participants, data were available on the EPQ-R (Eysenck et al., 1985; Ruch, 1999). Means and standard deviations on the four scales of the EPQ-R were M = 14.54, SD = 5.96 (Extraversion); M = 9.69, SD = 5.12 (Neuroticism); M = 8.00, SD = 4.61 (Psychoticism); and M = 6.89, SD = 5.02 (Lie). The estimates for internal consistency (α)/split-half reliability (r_{tt}, Spearman-Brown corrected) within the sample were .80/.83 (Extraversion), .81/.81 (Neuroticism), .46/.39 (Psychoticism), and .63/.56 (Lie).

Table 1 summarizes the data for three relevant covariates per participant group: Means and standard deviations for age, verbal IQ, and the Lie scale subtotal. According to separate one-way ANOVAs, the three groups of participants differed significantly in terms of age, $F(2, 161) = 7.33$, $p < .001$, and verbal IQ, $F(2, 126) = 7.59$, $p < .001$. As Bonferroni post hoc tests revealed, the child sexual abusers were on average significantly older than both the nonsexual offenders ($p = .002$) and the community controls ($p = .005$). On average, community controls achieved significantly higher verbal IQ scores than both child sexual abusers ($p = .037$) and nonsexual offenders ($p = .003$). The differences in the group means on the Lie scale of the EPQ-R approached

Table 2. Estimates of Internal Consistency and Reliability for Picture Categories Within Affinity in Two Tasks (Image Rating and Viewing Time).

	Image Rating		Viewing Time	
Picture Category	α	r_{tt}	α	r_{tt}
Female stimuli				
1. Adult	.92	.94	.86	.90
2. Juvenile	.90	.89	.79	.81
3. Prejuveniles	.96	.95	.85	.83
4. Small children	.95	.91	.88	.91
Male stimuli				
1. Adult	.97	.98	.83	.84
2. Juvenile	.96	.93	.89	.89
3. Prejuveniles	.98	.97	.83	.82
4. Small children	.96	.94	.82	.84

Note: r_{tt} = split-half correlation coefficient, Spearman-Brown corrected.

statistical significance, $F(2, 124) = 3.05$, $p = .051$, with child sexual abusers obtaining a higher group mean than community controls or nonsexual offenders.

Data on the SSPI (Seto & Lalumière, 2001) were only available for the child sexual abuser sample ($n = 42$) since the scoring of the SSPI requires case files on prior convictions for at least one sexual offense against a child. The mean on the SSPI was $M = 3.79$ ($SD = 1.16$, range: 1-5), with a median value of 4. Internal consistency of the four items of the SSPI was estimated at $\alpha = .07$ in the sample. The split-half correlation coefficient, corrected with the Spearman–Brown formula, was $r_{tt} = .28$.

Of all participants for whom data on sexual orientation based on two items from the SASO (Sell, 1996) were available ($n = 132$), about three quarters (78.0%) described themselves as exclusively heterosexual. About one seventh of participants indicated at least some homosexual interest (15.2%) and 6.8% indicated that they were asexual (neither hetero- nor homosexual). Across all three groups, 3.8% ($n = 5$) of participants indicated a stronger homo- than heterosexual orientation. Most of the participants who expressed a stronger homosexual orientation were from the child sexual abuser group ($n = 4$), compared to one individual from the community controls. The frequencies differed significantly across the three groups ($p = .004$ in a Fisher's exact test).

Reliability

Internal consistency estimates for the measurements within Affinity ranged from $\alpha = .79$ to $\alpha = .82$ for viewing times and from $\alpha = .90$ to $\alpha = .98$ for image ratings. Split-half reliability coefficients (corrected toward original length of scale using the Spearman–Brown formula) were estimated at $r_{tt} = .81$ to $r_{tt} = .91$ for the viewing times and $r_{tt} = .89$ to $r_{tt} = .98$ for the image ratings. Table 2 summarizes the results for all eight picture categories of age and gender.

Table 3. Analyses of Variance for Image Ratings (Top) and Viewing Time (Bottom).

Source	*df*	*F*	*P*	η_p^2
Image ratings				
Between subjects				
Group	2/161	4.23	.016	.05
Within subjects				
Picture category	2.28/367.10[a]	0.30	.770	.00
Group × Picture Category	4.56/367.10[a]	8.02	<.001	.09
Viewing time				
Between subjects				
Group	2/161	8.60	<.001	.10
Within subjects				
Picture category	2.06/332.35[a]	0.13	.881	.00
Group × Picture Category	4.13/332.35[a]	3.15	.014	.04

Note: The ANOVA (analysis of variance) are based on residuals, controlling for participants' age through linear regression. *df* = degrees of freedom between/within factors.
[a]Adjusted for unequal variances (Greenhouse–Geisser method). η^2_p = partial eta squared.

Validity

The subsequent analyses are based on the maximum of the median values for photographs of male or female individuals from the four age categories within Affinity (small children, prejuveniles, juveniles, and adults). Preliminary analyses revealed that the outcome of viewing time and image ratings could be predicted from participants' age. Hence, we separately regressed every viewing time and image rating variable on age, using only the unstandardized residuals from these analyses for the subsequent analyses in the remainder of the Results section. The results of the linear regression analyses are summarized in Table A1 in the Appendix. The correlation coefficients between image ratings and participants' age ranged from $r = .01$ (adults) to .20 (small children). With respect to viewing times, the correlation coefficients with participants' age were between $r = .21$ (adults) and .32 (small children).

Criterion validity. We conducted two separate ANOVAs to determine whether the three groups of participants differed with respect to overall image ratings and viewing times. For the *image ratings*, a 3 (group) × 4 (picture age categories) mixed-model ANOVA showed a significant main effect for Group and a significant Group by Age interaction (see Table 3, top). Similarly, a 3 (group) × 4 (picture age categories) mixed-model ANOVA for *viewing times* indicated a significant main effect for Group and a significant Group by Age interaction (see Table 3, bottom).

As a Tamhane-T2 post hoc test revealed, across all picture age categories the child sexual abusers differed significantly only from the community controls at $p = .003$, but not from the nonsexual offenders ($p = .182$) with regard to viewing time. A reverse pattern was observed for image ratings where a Tamhane-T2 test was significant with

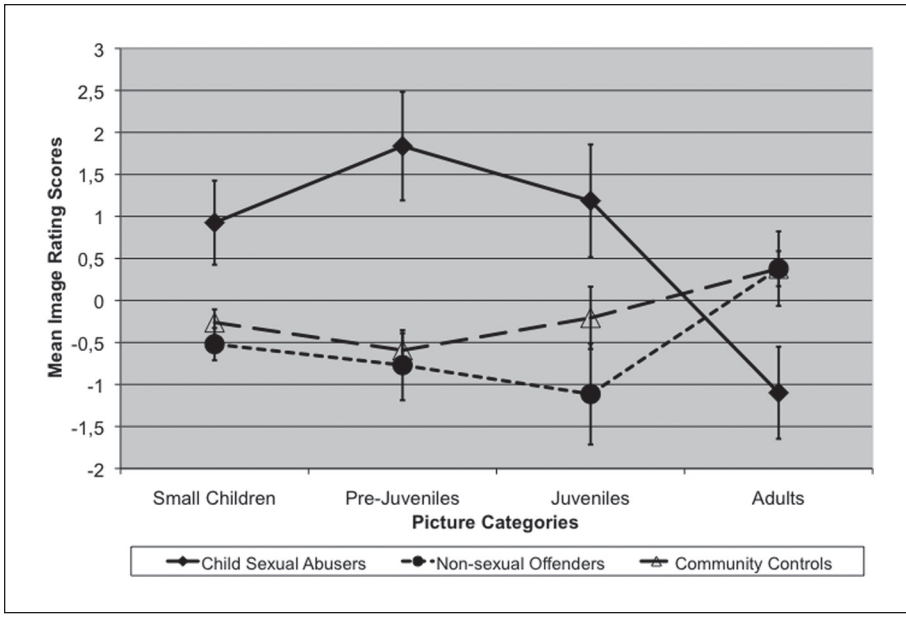

Figure 1. Mean Image Ratings per Picture Age Category for Three Groups of Participants: Child Sexual Abusers (n = 42), Nonsexual Offenders (n = 27), and Community Controls (n = 95). Values Represent Unstandardized Residuals Based on Median Values, Regressed on Participants' Age. For Each Participant, the Larger Median for Photographs of Male or Female Individuals Was Chosen Within Each Picture Age Category. Vertical Bars Represent Standard Errors of the Mean (*SE*).

respect to nonsexual offenders (p = .043), but not with regard to community controls (p = .133) across all picture age categories.

As separate repeated-measures ANOVA per group showed, the significant Group by Age Category interaction of image ratings was primarily due to the child sexual abusers and the community controls. Within both groups, the main effect of picture age category reached statistical significance: $F(2.11, 86.48) = 6.14$, p = .003, partial η^2 = .13 (for child sexual abusers) and $F(2.03, 190.74) = 3.79$, p = .024, partial η^2 = .04 (for community controls). Within the group of child sexual abusers, the differences between prejuvenile and adult (p = .018) and between juvenile and adult pictures (p = .027) reached statistical significance in a post hoc Bonferroni test. Among the community controls, the differences between small children and prejuvenile (p = .005) and between prejuvenile and adult pictures (p = .007) reached statistical significance in a post hoc Bonferroni test. Figure 1 shows the mean *image ratings* per picture age category for the three groups of participants.

An analogous group-specific repeated-measures ANOVA for viewing time indicated that the observed significant Group by Age Category interaction was mostly due to the child sexual abusers. Only within that subgroup did the main effect of picture

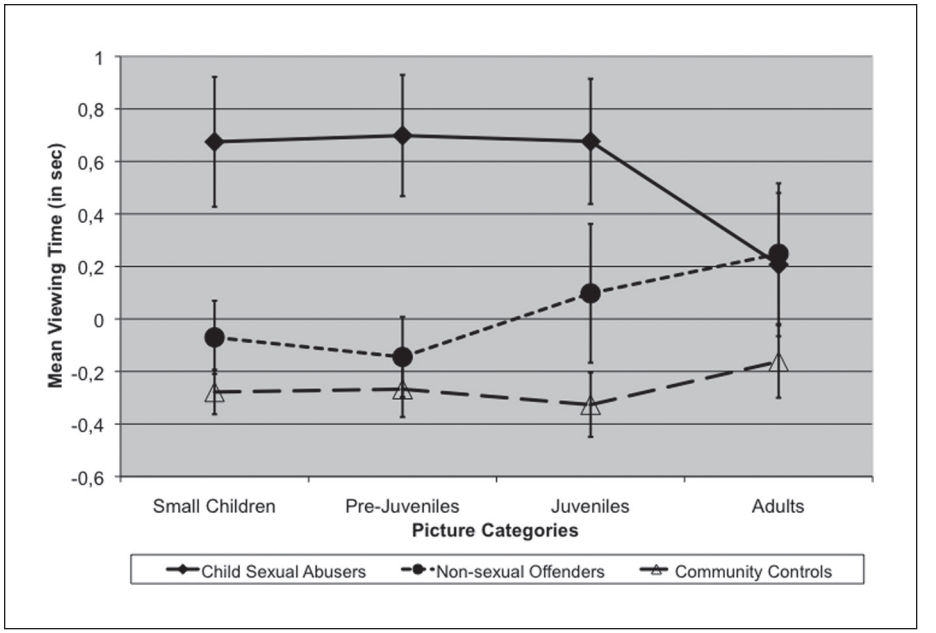

Figure 2. Mean Viewing Time (in Seconds) per Picture Age Category for Three Groups of Participants: Child Sexual Abusers (n = 42), Nonsexual Offenders (n = 27), and Community Controls (n = 95). Values Represent Unstandardized Residuals Based on Median Values, Regressed on Participants' Age. For Each Participant, the Larger Median for Photographs of Male or Female Individuals Was Chosen Within Each Picture Age Category. Vertical Bars Represent Standard Errors of the Mean (*SE*).

age category reach statistical significance: $F(2.12, 90.15) = 3.29$, $p = .037$, partial $\eta^2 = .08$ (with the differences between prejuvenile and adult, $p = .091$, and between juvenile and adult pictures, $p = .051$, approaching statistical significance in a post hoc Bonferroni test). Figure 2 shows the mean viewing times per picture age category for the three groups of participants.

As Figures 1 and 2 indicate, the group of child sexual abusers appeared distinct from the comparison groups of nonsexual offenders and community controls, both in terms of image ratings and in terms of viewing times.

We also conducted eight separate one-way ANOVAs, one for each of the four picture age categories between groups, for both image ratings and viewing times. As far as image ratings were concerned, the groups differed significantly on all four categories of pictures: adults, $F(2, 161) = 5.26$, $p = .006$; juveniles, $F(2, 161) = 3.44$, $p = .034$; prejuveniles, $F(2, 161) = 12.70$, $p < .001$; and small children, $F(2, 161) = 5.96$, $p = .003$. Tamhane-T2 post hoc tests showed that child sexual abusers gave significantly lower average ratings for images of adults ($M = -1.10$, $SD = 3.55$) than did community controls ($M = 0.38$, $SD = 2.03$, $p = .044$, $d = 0.57$). Conversely, child sexual abusers had significantly higher mean ratings for pictures of juveniles ($M = 1.19$, $SD = 4.35$) than

nonsexual offenders ($M = –1.11$, $SD = 3.13$) at $p = .039$ ($d = 0.60$). Concerning pictures of prejuveniles, child sexual offenders had significantly higher average ratings ($M = 1.84$, $SD = 4.19$) than both nonsexual offenders ($M = –0.77$, $SD = 2.17$, $p = .004$, $d = 0.75$) and community controls ($M = –0.59$, $SD = 1.98$, $p = .002$, $d = 0.86$). Finally, child sexual abusers gave significantly higher ratings for pictures of small children ($M = 0.93$, $SD = 3.24$), compared with the average ratings given by nonsexual offenders ($M = –0.52$, $SD = 1.00$, $p = .014$, $d = 0.56$) and community controls ($M = –0.26$, $SD = 1.51$, $p = .006$, $d = 0.55$).

For viewing time, the groups differed significantly on pictures of juveniles, $F(2, 161) = 8.47$, $p < .001$; prejuveniles, $F(2, 161) = 10.78$, $p < .001$; and small children, $F(2, 161) = 11.71$, $p < .001$, but not on pictures of adults. Tamhane-T2 post hoc tests indicated that for pictures of juveniles, child sexual abusers had significantly longer average viewing times ($M = 0.68$, $SD = 1.55$) than community controls ($M = –0.33$, $SD = 1.20$, $p = .001$, $d = 0.77$), for pictures of prejuveniles child sexual abusers had significantly longer average viewing times ($M = 0.70$, $SD = 1.50$) than nonsexual offenders ($M = –0.14$, $SD = 0.79$, $p = .009$, $d = 0.67$) and community controls ($M = –0.27$, $SD = 1.03$, $p < .001$, $d = 0.82$); and that for pictures of small children child sexual abusers ($M = 0.67$, $SD = 1.60$) had significantly longer average viewing times than nonsexual offenders ($M = –0.07$, $SD = 0.72$, $p < .016$, $d = 0.57$) and community controls ($M = –0.28$, $SD = 0.82$, $p < .001$, $d = 0.86$). The significant Group × Picture Age Category interactions in the mixed ANOVA as well as the significantly higher image ratings and longer viewing times with respect to the photographs of small children in the follow-up one-way ANOVA were as expected.

Concurrent validity. Using the raw data of the entire sample we calculated the correlations between image ratings and viewing times for the picture categories as defined by age and gender of the persons depicted in the photographs. For the photographs of female individuals, the correlation coefficients between median ratings and median viewing times of the participants were $r = .29$ (small children), .30 (prejuveniles), .23 (juveniles), and .00 (adults). For the photographs of male individuals, the correlation coefficients between median ratings and viewing times of the participants were $r = .21$ (small children), .34 (prejuveniles), .38 (juveniles), and .39 (adults). All correlation coefficients (except the one concerning pictures of adult women) reached statistical significance in one-sided tests at an overall Type 1 error level of $p < .05$, using the Šidák correction for a total of eight comparisons ($p < .006$). This outcome was as expected in our hypothesis.

Based on the unstandardized residuals (see introduction to the section Validity above) we calculated differential indices of pedophilic interest for both image ratings and viewing times. For this purpose, we subtracted the higher value for pictures showing adults from the corresponding value for pictures of small children. In the subsample of child sexual abusers ($n = 42$) the pedophilic interest index based on image ratings did not correlate with the total score from the Screening Scale of Pedophilic Interests (SSPI; Seto & Lalumière, 2001; $r = .03$, $p = .42$, *ns*, one-sided). The corresponding differential index of pedophilic interest based on viewing time did not correlate with the SSPI total score either ($r = .05$, $p = .37$, *ns*, one-sided). The same pattern

Table 4. Correlations (and Corresponding *p* Values) of Verbal IQ ($n = 129$) and the Lie Scale ($n = 127$) With Image Ratings and Viewing Time for Four Categories of Pictures.

	Image Rating		Viewing Time	
Picture Category	IQ	Lie	IQ	Lie
1. Adult	.08	–.02	.04	–.05
	(.39)	(.86)	(.66)	(.57)
2. Juvenile	–.14	.15	–.10	–.07
	(.13)	(.09)	(.25)	(.44)
3. Prejuvenile	–.09	.11	–.05	–.03
	(.32)	(.24)	(.58)	(.75)
4. Small children	–.17	.14	–.08	–.02
	(.06)	(.13)	(.38)	(.81)

Note: Verbal IQ was measured with the MWT-B Multiple Choice Vocabulary Test (Lehrl, 1999; see Lehrl, Triebig, & Fischer, 1995, 1995). Socially desirable responding was assessed with the Lie scale of the Eysenck Personality Questionnaire—Revised (EPQ-R; Eysenck, Eysenck, & Barrett, 1985; Ruch, 1999). Correlation coefficients are Pearson product–moment correlation coefficients with *p* values from two-sided testing.

held if the differences between the prejuvenile and adult picture categories were analyzed: $r = .11$ ($p = .24$, *ns*, one-sided) for image ratings and $r = .15$ ($p = .17$, *ns*, one-sided) for viewing time. This outcome, a nonsignificant linear relationship of both kinds of indices (image ratings and viewing time) for pedophilic interest with the SSPI total score, was contrary to our hypothesis.

Discriminant validity. On the basis of raw scores and using the maximum of the medians for either male or female pictures per age category, we calculated the correlations with verbal intelligence as measured with the MWT-B (Lehrl, 1999) and with the Lie scale of the EPQ-R (Eysenck et al., 1985; Ruch, 1999). Table 4 summarizes these results.

As far as verbal IQ was concerned, the correlations were small and statistically nonsignificant (all $p > .05$), with average correlations of –.08 (image ratings) and –.05 (viewing time). Similarly, the correlations with respect to the Lie scale remained nonsignificant (all $p > .05$). The average correlation of the Lie scale with image ratings was .10 and –.04 with viewing time. The lack of any significant correlation of either image ratings or viewing times with verbal intelligence or a socially desirable response style was in accordance with our hypothesis.

Classification Accuracy

Three indices of pedophilic interest were calculated, based on the unstandardized residuals (controlling for participants' age): (a) a differential index based on viewing times, (b) a differential index based on image ratings, and (c) a posterior probability derived from a multivariate normal Bayesian classification algorithm. The differential indices were calculated by subtracting the values derived from the maximum

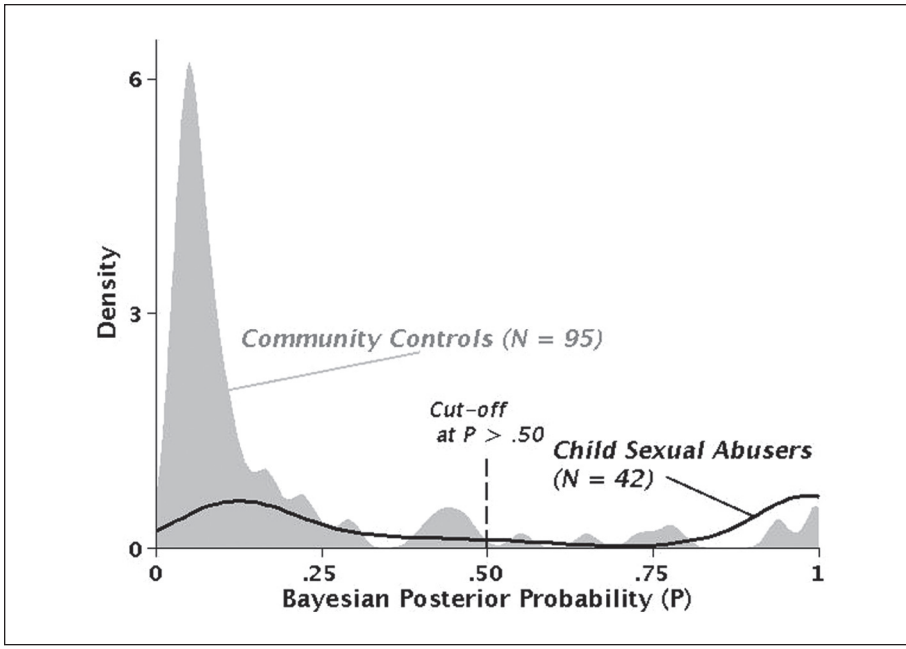

Figure 3. Kernel-Density Estimates for the Distributions of Bayesian Posterior Probabilities Among Community Controls and Child Sexual Abusers. Density Curves Were Scaled According to Relative Group Size.

median score for any adult picture category (male or female) from the corresponding score derived from the maximum median of any small children picture category (boy or girl).

Comparing the two groups of child sexual abusers and community controls, the groups differed significantly on both accounts. Child sexual abusers had significantly higher average scores on the differential index based on image ratings ($M = 2.02$, $SD = 5.31$) than community controls ($M = –0.64$, $SD = 2.62$): $t(135) = 3.94$, p (one-sided) $< .001$ ($d = 0.73$). The child sexual abusers also had significantly higher average scores on the differential index based on viewing times ($M = 0.47$, $SD = 1.51$), compared with the community controls ($M = –0.12$, $SD = 1.22$): $t(135) = 2.40$, p (one-sided) $= .009$ ($d = 0.43$).

For the multivariate classification algorithm, the optimal combination of variables was identified among all 255 possible combinations, derived from the set of 2 (image rating, viewing time) × 4 (picture age categories) = 8 variables. The optimal combination entailed the following four variables: viewing time (adult images), viewing time (juvenile images), viewing time (prejuvenile images), and image rating (prejuvenile images). In a cross-validation by the leave-one-out method, every case was assigned

based on a classification routine derived from all other cases. Each case was assigned to that group for which it achieved the highest posterior probability. Figure 3 plots the distribution of these probabilities separately for the individuals from the child sexual abuser group and for the community controls.

With a mean value of $M = 0.53$ ($SD = 0.40$), the child sexual abusers on average had a significantly higher posterior probability than the community controls ($M = 0.19$, $SD = 0.25$): $t(135) = 6.06$, p (one-sided) $< .001$ ($d = 1.13$). The fact that the child sexual abusers had significantly higher values on all three indices of pedophilic interest (i.e., on the two univariate differential indices as well as on the multivariate one) was according to expectation.

Next, we compared the diagnostic accuracy of the three indices of pedophilic interest. The effect size for the differentiation between child sexual abusers and community controls, measured as the area under a receiver operating characteristic curve (ROC-AUC), was AUC = .62 ($p = .026$, 95% CI: [.51-.72]) for the differential index based on viewing time, AUC = .64 ($p = .008$, 95%-CI: [.54-.74]) for the differential index based on image ratings, and AUC = .80 ($p < .001$, 95% CI: [.72-.89]) for the multivariate normal Bayesian classification based on four individual criteria (i.e., viewing time for adult, juvenile, and prejuvenile images as well as image ratings for prejuveniles). According to the method by DeLong et al. (1988) the AUC value for the multivariate normal Bayesian classification procedure was significantly higher (with $p = .004$) than the next largest AUC of the preferential index based on image ratings. This outcome supports our hypothesis that a multivariate index would be more accurate than a univariate one. The multivariate index differentiated equally well for child molesters with girl ($n = 16$) and with boy victims ($n = 25$), with AUC values of .836 ($SE = .047$, $p < .001$) and .784 ($SE = .049$, $p < .001$), respectively.

For identifying more than 90% of child sexual abusers correctly with the Bayesian posterior probabilities obtained from the multivariate classification algorithm (sensitivity: 90.5%), one would face a rather low specificity of 50.5%. In other words, for correctly recognizing more than 9 out of 10 child sexual abusers (true-positive cases), about half of the community controls would be wrongly assumed to be child sexual abusers (false-positive cases). To achieve similar levels of sensitivity (90%) with the univariate differential preference indices based on viewing times or image ratings, one would obtain even lower specificities of 13.7% and 7.4%, respectively.

Setting the threshold for sensitivity to 50% (i.e., at a margin where half of the child sexual abusers were assigned correctly), the specificity values were 62.1% (viewing time preference index), 82.1% (image rating preference index), and 86.3% (multivariate Bayesian classification). For the multivariate classifier, the overall rate of correct classification would be 75% at this threshold.

Discussion

We examined the reliability, validity, and diagnostic properties of a computerized assessment tool for sexual preference (Affinity 2.5). The Affinity procedure is based on explicit ratings of pictures of individuals from either gender and from four age

categories (small children to adults) as sexually attractive as well as on the unobtrusively measured viewing time that testees take for this task. Participants were pedophilic child sexual abusers who acknowledged their offenses (i.e., individuals who admitted to sexual offenses against children but not necessarily to pedophilic interest), nonsexual offenders, and community controls, all of them male adults.

In keeping with earlier assessments of a prior version of the Affinity software with adult (Croxen, 2003) and adolescent participants (Worling, 2006), both internal consistency and reliability estimates were in a range (> .80) that is commonly deemed as sufficient for research purposes, but insufficient (≤ .90) for single-case diagnostics (see Nunnally & Bernstein, 1994), at least as far as most of the viewing time variables are concerned. Consequently, it may be worthwhile to explore the possibility of an optimal weighting of individual test scores within a test battery for actual applications by Green's (1950) method, for instance.

Criterion validity was tested through two ANOVAs, one each for image ratings and viewing times across the four picture age categories within Affinity. The expected significant Group × Picture Age Category interaction effect (with child sexual abusers having higher mean scores on images of small children and prejuveniles) was observed for both criteria (image ratings and viewing times). The effect size was higher for the former than for the latter, however: The partial η^2 of .09 signified a medium effect for the interaction effect involving image ratings, and the partial η^2 of .04 indicated a small effect for the corresponding interaction effect based on viewing time.

Comparing the evaluations of pictures and the corresponding viewing times separately for images of small children, prejuveniles, juveniles, and adults, child sexual abusers had higher mean values for pictures of both small children and prejuveniles, an outcome that was expected and that appears plausible in the light of previous research combining the measurement of viewing time for photographs with self-reports of sexual interest among admitting child molesters (e.g., Abel et al., 1998, 2001; Banse et al., 2010; Croxen, 2003; Gress, 2005; Worling, 2006). The average score difference between child sexual abusers and either of the other two groups for photographs of small children and prejuveniles reflected medium to large effects, with *d* scores ranging from 0.55 to 0.86 for viewing times and image ratings. It is noteworthy that we controlled for participants' age through linear regression, using only the remaining unstandardized residuals for the analyses since participants' age correlated significantly with image ratings and viewing times for nearly all picture age categories (except for the ratings of adult images).

Under the ANOVA framework we opted against the use of ipsative measures. Ipsatization through individual *z* scores could compromise the data for comparisons across individuals (see Cheung, 2006). Harris et al. (1992, p. 504) noted, "Of course, such a conversion to ipsative measures obviates comparison between groups on overall response magnitude." In applied clinical circumstances, *z*-score ipsatization may yield crucial information in a single-case diagnostic framework and is therefore part of the analysis routine within Affinity.

Internal concurrent validity of measurements within Affinity was ascertained at a medium effect size for most instances, with image ratings and viewing times correlating significantly with each other at between $r = .21$ and .39 across picture age categories. The sole exception was the category of images of adult women, with a zero correlation between ratings and viewing times. This latter result mirrors prior results with Affinity (Worling, 2006). Although we did not hypothesize this outcome, Worling's explanation may be correct that this has to do with exaggerated ratings for women in an effort to appear socially acceptable: Post hoc analyses revealed that the mean ratings for adult women were significantly correlated with the subtotal on the Lie scale at $r = .52$ ($p = .016$, two-sided) among the child sexual abusers only, but not among community controls ($r = -.08$, $p = .44$) or nonsexual offenders ($r = -.49$, $p = .052$). The group of child sexual abusers had the highest mean on the Lie scale. The mean differences toward the other two groups of participants did not reach statistical significance, however (see Table 1).

External concurrent validity could not be ascertained: Contrary to expectation neither viewing time- nor rating-based indices of pedophilic preference correlated with the total score from a brief actuarial measure of pedophilic interest based on the sexual offense history (SSPI; Seto & Lalumière, 2001). Although the present sample of child sexual abusers achieved a higher mean score and lower variance than the corresponding samples in previous studies that utilized the SSPI (e.g., Mokros, Dombert et al., 2010; Seto & Lalumière, 2001; Seto et al., 2004), the standard deviation was on a par with the sample from Banse et al. (2010). Banse et al. did observe significant correlations of the SSPI total score with external variables (such as sexual interest in male children, assessed via self-report and viewing time). Still, the reliability estimates of the SSPI were low both in the present sample and in the sample assessed by Banse et al. (2010), with $r_{tt} = .28$ and $\geq .38$, respectively. Hence, other variables for external validation than the four-item SSPI scale or an extended (and more reliable) version of the SSPI may prove useful in future studies.

Comparisons with a scale that measures socially desirable responding (the L scale of the EPQ-R; Eysenck et al., 1985) as well as with a measure of verbal intelligence (Lehrl, 1999) indicated independence for the measures within Affinity. With due caution, this outcome can be interpreted as indicative of discriminant validity for both image ratings and viewing times. It may be advisable, however, to incorporate a measure of socially desirable responding directly into the viewing time methodology in future studies. In the present study, the tendency toward faking good (in terms of a socially desirable response style) was assessed through a subscale of a self-report personality questionnaire. Apart from these conceptual differences, the Lie scale of the EPQ-R may not have been sensitive enough to register tendencies for faking good in the present experimental design given its limited scale reliability of $r_{tt} = .56$.

Simple differential indices of pedophilic preference (that compared the values for the small children picture category with the values for the adult category) yielded a discriminatory power that was above chance expectation yet would entail considerable rates of false positive decisions. In this regard, the performance of such differential

indices within Affinity was clearly below the rates that other researchers reported with similar methods (e.g., Abel et al., 1998; Banse et al., 2010; Mokros, Dombert et al., 2010) but also lower than prior unpublished research of Affinity with adult sexual offenders indicated (Croxen, 2003). The 95% confidence intervals of the AUCs for both viewing time (.62) and image ratings (.64) found in the present study overlap with the ones reported by Worling (2006) for adolescent sexual offenders with regard to child victims: Worling noted AUC values (with 95% confidence intervals) of .61 [.47-.75] for viewing times and of .67 [.54-.79] for image ratings.

In terms of both viewing times and image ratings, the child sexual abusers showed a distinct profile compared with nonsexual offenders and community controls. A multivariate classification algorithm (based on four individual variables: viewing times for adults, juveniles, and prejuveniles, and image ratings of prejuveniles) yielded a classification accuracy (AUC = .80) that surpassed the results from the univariate indices to a significant degree. This outcome is plausible in light of the earlier findings on the incremental validity of composite scores within Affinity by Worling (2006). Moreover, this result confirms the supposition by Harris et al. (1996) that combined indices that include viewing time and self-report data on sexual interest in children might prove more useful than single variables. Similarly, Laws et al. (2000) pointed out the potential usefulness of a battery of tests pertaining to deviant sexual interest. Still, adopting a *more likely than not* criterion as a rule for allocation, the rate of false positives in the present sample was substantial (>10%), even under the multivariate classification method. The rate of correct classification (75%) was lower than in other comparable studies (Banse et al., 2010: 90%; Mokros, Dombert et al., 2010: 98%), but on a par with the rate reported by Abel et al. (2001: 78%) who tested solely child molesters who explicitly denied their offenses, however. Likewise, Gress (2005) found an overall rate of correct classification of 77% with viewing time methodology. Nearly all of the offenders in the study by Gress (2005) admitted their offenses. It should be noted, however, that the overall rate of correct classification not only depends on the diagnostic utility of the assessment method but also on the relative proportions of child sexual abusers and controls in the sample (as well as on the chosen cutoff). Although the multivariate classification routine involved a leave-one-out cross-validation procedure, there is still a considerable chance of overfitting. Therefore, the results remain tentative until the combination of predictor variables from Affinity has been tested in other samples.

The prior probability that informed the multivariate classification routine was estimated from the sample (here: 30.7%). That is, all child sexual abusers were regarded as pedophilic (which seems tenable given that they had been diagnosed clinically as such), whereas all controls were deemed nonpedophilic. The value of 30.7% falls within the margin of 25% to 40% that was estimated as the proportion of actual pedophiles among men convicted of child sexual abuse (Schaefer et al., 2010; Seto, 2008). Hence, the margin may be somehow comparable to diagnostic situations within forensic settings (i.e., in situations where evidence of committing child sexual abuse is at

hand). Since the prevalence of pedophilic interest among men in the community was estimated at between 0.2% and 3.8% in surveys (Ahlers et al., 2011; Santtila et al., under review), the application of Affinity as a screening tool would likely produce very large numbers of false positive cases. Meehl and Rosen (1955) provide a formula that informs about the degrees of sensitivity and specificity that a test would need to achieve to produce more true-positive than false-positive results, given a particular prevalence. If the prevalence of pedophilia in the male population was as high as 3%, a test with a sensitivity of 97% or higher would have to achieve a specificity of more than 97%. Only then it would be slightly more likely than not that any case diagnosed as positive would in fact be pedophilic. Therefore, readers should be reminded that Affinity was not developed as a screening device but as a complementary aid to help investigating the sexual interests of men (Glasgow, 2009). The program should be part of a comprehensive assessment and not be used in isolation.

But also before Affinity could be used for forensic diagnosing and decision making, the issue of fakability warrants closer scrutiny. It is a shortcoming of the present study that it neither addressed the performance of Affinity with explicit deniers nor the possibility of faking results by motivated participants. All child molesters in the present sample were diagnosed with pedophilia according to the ICD-10 criteria. They at least acknowledged their offense(s), even though not all of them may have admitted to sexual interest in children. Future studies should explicitly ask participants whether they felt a sexual inclination toward children, possibly using the questionnaires by Croxen (2003) or Banse et al. (2010). Before such endeavors are undertaken, though, it seems worthwhile to consider the use of larger and possibly more sexually relevant images within Affinity, such as the Not Real People Set (NRP; Pacific Psychological Assessment Corporation, Inc., 2004) or the Virtual People Set (Mokros, 2009). The NRP has been demonstrated to be a reliable set of stimuli (Laws & Gress, 2004; Mokros et al., 2011) that has been used successfully in a range of studies on implicit measures of pedophilic sexual interest, some of them affording higher criterion validity than the present study (Banse et al., 2010; Mokros, Dombert et al., 2010). In future studies of Affinity with deniers or instructed fakers, attention should also be paid to contradictory profiles on explicit (image rating) and implicit (viewing time) measures, possibly incorporating the additional ranking task of manikins that represent differences in age and gender (Glasgow, 2009).

Given the size of the sample of child sexual abusers ($n = 42$) it is conceivable that some of the results were due to specific properties of the individuals taking part in the study rather than reflecting the general population of child sexual abusers. Conceptually, the present study followed the quasi-experimental *known group* approach (Banse, Seise, & Zerbes, 2001; see also De Houwer et al., 2009): applying the assessment method in a group that presumably varies with regard to the attribute that is to be measured. As the individuals in the group of child sexual abusers had been diagnosed with pedophilia and obtained relatively high scores on the Screening Scale of Pedophilic Interests (SSPI; Seto & Lalumière, 2001), they most likely represented a suitable

reference group. On the other side it was not possible to ascertain for the individuals from the nonsexual offender and community samples whether they did in fact not have any sexual interest in children. The prevalence estimates mentioned above (Ahlers et al., 2011; Santtila et al., in press) offer some guidance, with self-expressed sexual interest in children present among up to 3.8% of respondents. Hence, it is conceivable that at least some of the individuals from the nonsexual offender and community samples may have had a sexual interest in children, thus possibly inflating the observed false-positive rates.

Furthermore, the child sexual abusers were tested at various stages of the treatment process. It would have been beneficial to either test all participants before the onset of treatment or only to choose participants who were at comparable stages of the treatment process. Consequently, we cannot exclude the possibility that therapy progress may have confounded the responses in the Affinity testing.

Another limitation is the low reliability of the SSPI in the present sample of child sexual abusers. The lack of concurrent validity with the Affinity parameters may therefore be due to the low reliability of the SSPI rather than due to the Affinity measures. Hence, it would be sensible to use other established indicators of pedophilic sexual interest (such as deviant sexual arousal) for comparison purposes in future studies.

In sum, the present findings add to the knowledge that the processing of child-related stimuli (here: pictures) is altered in pedophilic men. In the long run, this processing bias may be used for diagnostic purposes in clinical and forensic settings. Given the present finding on the reliability and validity of the Affinity software self-report and viewing time methods should be further refined for their inclusion into comprehensive assessments of pedophilia.

Table A1. Univariate Linear Regressions of Median Image Ratings (Left) and Median Viewing Times (Right) on the Age of the Participants.

	Image Rating			Viewing Time		
Picture Category	*a*	*b*	*r*	*a*	*b*	*r*
1. Adult	2.912	0.001	.01	2.677	0.026	.21 **
2. Juvenile	–4.633	0.060	.19 *	2.116	0.036	.31 ***
3. Prejuvenile	–6.945	0.047	.19 *	1.283	0.032	.31 ***
4. Small children	–7.176	0.034	.20 *	1.171	0.030	.32 ***

Note: *a* = intercept parameter; *b* = slope parameter; *r* = correlation coefficient.
*$p < .05$. **$p < .01$. ***$p < .001$.

Appendix

Table A1 summarizes the slope (b) and intercept parameters (a) of the regression equations $y = bx + a$ for median *viewing times* (x) and for median *image ratings* across the four picture age categories.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: David Glasgow, Richard Laws, and Carmen Gress have a financial interest in the Not Real People (NRP) stimulus set and the *Affinity 2.5* assessment procedure.

Notes

1. Generally, the notion of a measure would imply that the attribute in question does exist and that changes in the attribute would lead to noticeable changes in the outcome (Borsboom, Mellenbergh, & van Heerden, 2004). More specifically, for a measure to be regarded as implicit (De Houwer et al., 2009) (a) the attribute that causes the measurement outcome has be known, (b) the process by which the attribute causes the measurement outcome must be understood, and (c) the attribute should cause the measurement outcome in an automatic way. If the link between the process and the measurement outcome was present even in unaware, unmotivated, distracted, or time-pressured subjects, this might be regarded as indicative of automaticity.
2. Since the area under a receiver operating characteristic curve (ROC-AUC) was not formally introduced within this article yet, the AUC value of .87 reported by Croxen (2003) was converted into a d score using the formula given in Ruscio (2008), assuming equal variances for both groups: $d = \sqrt{2} \times F^{-1}(AUC)$, where F^{-1} is the inverse of the standard normal cumulative distribution function (see Rice & Harris, 2005, p. 619; Swets, 1986, eq. 21).
3. In a bootstrap reanalysis (with 10^6 draws) of the data reported by Croxen (2003), the 99% confidence interval of the hit rate expected by chance was [33.9%-66.1%].
4. For all but one participant from the child sexual abuser group, age of the youngest victim was 13 years or lower. One participant with an older victim (age: 16 years) was included since he reported being sexually aroused by prepubescent children.
5. This follows from (a) definition of the reliability of a test X, $\rho_{XX'}$, as the squared correlation between the observed scores on said test (X) and the true scores (T), ρ^2_{XT}: $\rho^2_{XT} = \rho_{XX'}$; (b) the identity $\rho_{XY} = \rho(X,T_Y)\rho(Y,T_Y) \Leftrightarrow \rho(X,T_Y) = \rho_{XY} / \rho(Y,T_Y) = \rho_{XY} / \sqrt{\rho_{YY'}}$ (eq. 3.9.2 in Lord & Novick, 1968, p. 70); and (c) the inequality $\rho(X,T_Y) \leq \rho(X,T_X)$ (eq. 3.9.1b in Lord & Novick, 1968, p. 69). Consequently, the square of $\rho(X,T_Y)$, $\rho(X,T_Y)^2$, yields a lower bound estimate for ρ^2_{XT} or $\rho_{XX'}$: $\rho_{XX'} = \rho(X,T_X)^2 \geq \rho(X,T_Y)^2 = \left(\rho_{XY} / \sqrt{\rho_{YY'}}\right)^2$. Applied to the data reported

for the SSPI by Banse et al. (2010), with the lower bound of the reliability for the external variable given by Cronbach's $\alpha = .88$, this yields: $r_{tt} \geq [.58/.88^{(1/2)}]^2 = .38$.

References

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., Hand, C. G., Holland, L. A., & Phipps, A. (2001). Classification models of child molesters utilizing the Abel Assessment for Sexual Interest™. *Child Abuse & Neglect, 25*, 703-718. doi:10.1016/S0145-2134(01)00227-7

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

Abel Screening, Inc. (n.d.). *The Abel Assessment for Sexual Interest–2*™ (AASI-2). Retrieved from http://www.abelscreen.com/tests/aasi.html

Ahlers, C. J., Schaefer, G. A., Mundt, I. A., Roll, S., Englert, H., Willich, S. N., & Beier, K. M. (2011). How unusual are the contents of paraphilias? Paraphilia-associated sexual arousal patterns in a community-based sample of men. *Journal of Sexual Medicine, 8*, 1362-1370. doi:10.1111/j.1743-6109.2009.01597.x

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

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes toward homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48*, 145-160.

Beech, A. R., Kalmus, E., Tipper, S. P., Baudouin, J.-Y., Flak, V., & Humphreys, G. W. (2008). Children induce an enhanced attentional blink in child molesters. *Psychological Assessment, 20*, 397-402. doi:10.1037/a0013587

Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment, 13*, 118-126. doi:10.1037/1040-3590.13.1.118

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071. doi:10.1037/0033-295X.111.4.1061

Brown, A. J. (2005). *Ipsative score distortion on Affinity 2.0*. Unpublished master's thesis, Brigham Young University, Provo, UT.

Cantor, J. M., Blanchard, R., Robichaud, L. K., & Christensen, B. K. (2005). Quantitative reanalysis of aggregate data on IQ in sexual offenders. *Psychological Bulletin, 131*, 555-568. doi:10.1037/0033-2909.131.4.555

Cheung, M. W. L. (2006). Recovering preipsative information from additive ipsatized data: A factor score approach. *Educational and Psychological Measurement, 66*, 565-588. doi:10.1177/0013164405282486

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi:10.1037/0033-2909.112.1.155

Croxen, J. (2003). *The relationship between viewing time and sexual interest in paedophilic and normal men: The utility of a new computer assisted assessment too*l. Unpublished doctoral dissertation, University of Manchester, Manchester, UK.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347-368. doi:10.1037/a0014211

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*, 837-845. doi:10.2307/2531595

Eysenck, H. J., Eysenck, S. B. G., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences, 6*, 21-29. doi:10.1016/0191-8869(85)90026-1

Geer, J. H., & Melton, J. S. (1997). Sexual content-induced delay with double-entendre words. *Archives of Sexual Behavior, 26*, 295-316. doi:10.1023/A:1024574915201

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

Giotakos, O. (2006). A combination of viewing reaction time and incidental learning task in child molesters, rapists, and control males and females. *Annals of General Psychiatry, 5*(Suppl. 1), S114. doi:10.1186/1744-859X-5-S1-S114

Glasgow, D. V. (2009). Affinity: The development of a self-report assessment 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.

Glasgow, D. V., Osborne, A., & Croxen, J. (2003). An assessment tool for investigating paedophile sexual interest using viewing time: An application of single case methodology. *British Journal of Learning Disability, 31*, 96-102. doi:10.1046/j.1468-3156.2003.00180.x

Gray, N. S., Brown, A. S., MacCulloch, M. J., Smith, J., & Snowden, R. J. (2005). An implicit test of the associations between children and sex in pedophiles. *Journal of Abnormal Psychology, 114*, 304-308. doi:10.1037/0021-843X.114.2.304

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 the Law, 1*, 1-10.

Green, B. F., Jr. (1950). A note on the calculation of weights for maximum battery reliability. *Psychometrika, 15*, 57-61. doi:10.1007/BF02289178

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. (2007). *Delays in attentional processing when viewing sexual imagery: The development and comparison of two measures*. Unpublished dissertation, University of Victoria, Victoria, British Columbia, Canada.

Gress, C. L. Z., & Laws, D. R. (2009). Measuring sexual deviance: Attention-based measures. In A. R. Beech, L. A. Craig, & K. D. Browne (Eds.), *Assessment and treatment of sexual offenders: A handbook* (pp. 109-128). Chichester, UK: Wiley.

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. *Behavioural 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

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

Janssen, E., Everaerd, W., Spiering, M., & Janssen, J. (2000). Automatic processes and the appraisal of sexual stimuli: Toward an information processing model of sexual arousal. *Journal of Sex Research, 37*, 8-23. doi:10.1080/00224490009552016

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2005). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences, 103*, 17048-17052. doi:10.1073/pnas.0605678103

Kahneman, D. (1973). *Attention and effort*. Upper Saddle River, NJ: Prentice Hall.

Laws, D. R., & Gress, C. L. Z. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9*, 183-196. doi:10.1348/1355325041719338

Laws, D. R., Hanson, R. K., Osborn, C. A., & Greenbaum, P. E. (2000). Classification of child molesters by plethysmographic assessment of sexual arousal and a self-report measure of sexual preference. *Journal of Interpersonal Violence, 15*, 1297-1312. doi:10.1177/088626000015012004

Lehrl, S. (1999). *Mehrfachwahlwortschatz-Intelligenztest: MWT-B* [Multiple choice vocabulary intelligence test] (2nd ed.). Balingen, Germany: Spitta.

Lehrl, S., Triebig, B., & Fischer, B. (1995). Multiple Choice Vocabulary Test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica, 91*, 335-345. doi:10.1111/j.1600-0404.1995.tb07018.x

Letourneau, E. L. (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

Limestone Technologies, Inc. (n.d.). *Visual Sexual Preference Assessment*. Retrieved from http://www.limestonetech.com/index.cfm/product-lines/sex-offender-management/visual-sexual-preference-assessment/packages/

Loewinger Cloyd, L. (2007). *Relationship between key variables in penile plethysmograph and viewing time measures of sexual arousal in sex offending adult males.* Unpublished doctoral dissertation, University of Miami, Coral Gables.

Lord, F. M., & Novick, M. R. (Eds.) (1968). *Statistical theories of mental test scores, with contributions by Allan Birnbaum*. Reading, MA: Addison-Wesley.

Maner, J. K., Gailliot, M. T., Rouby, D. A., & Miller, S. L. (2007). Can't take my eyes off you: Attentional adhesion to mates and rivals. *Journal of Personality and Social Psychology, 93*, 389-401. doi:10.1037/0022-3514.93.3.389

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194-216. doi:10.1037/h0048070

Mihailides, S., Devilly, G. J., & Ward, T. (2004). Implicit cognitive distortions and sexual offending. *Sexual Abuse: A Journal of Research and Treatment, 16*, 333-350. doi:10.1177/107906320401600406

Mokros, A. (2009, October). *A computer-generated picture set for assessing pedophilic sexual interest*. Paper presented at the Association for the Treatment of Sexual Abusers (ATSA) 27th Annual Research and Treatment Conference, Dallas, TX.

Mokros, A., Butz, M., Dombert, B., Santtila, P., Bäuml, K.-H., & Osterheider, M. (2011). Judgment of age and attractiveness in a paired comparison task: Testing a picture set developed for diagnosing paedophilia. *Legal and Criminological Psychology, 16*, 323-334. doi:10.1348/135532510X514104

Mokros, A., Dombert, B., Osterheider, M., Zappalà, A., & Santtila, P. (2010). Assessment of pedophilic sexual interest with an attentional choice reaction time task. *Archives of Sexual Behavior, 39*, 1081-1090. doi:0.1007/s10508-009-9530-6

Mokros, A., Stadtland, C., Osterheider, M., & Nedopil, N. (2010). Assessment of risk for violent recidivism through multivariate Bayesian classification. *Psychology, Public Policy, and Law, 16*, 418-450. doi:10.1037/a0021312

Nunes, K. L., Firestone, P., & Baldwin, M. W. (2007). Indirect assessment of cognitions of child sexual abusers with the Implicit Association Test. *Criminal Justice and Behavior, 34*, 454-475. doi:10.1177/0093854806291703

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Ó Ciardha, C., & Gormley, M. (2012). Using a pictorial-modified Stroop task to explore the sexual interests of sexual offenders against children. *Sexual Abuse: A Journal of Research and Treatment, 24*, 175-197. doi:10.1177/1079063211407079

Pacific Psychological Assessment Cooperation, Inc. (2004). *The Not-Real People stimulus set for assessment of sexual interest*. Victoria, British Columbia, Canada: Author.

Pacific Psychological Assessment Corporation, Inc. (2007). The Affinity Measure of Sexual Interest—Version 2.5 [Computer Software]. Retrieved from http://pacific-psych.com/products/affinity.html

Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Poeppl, T. B., Nitschke, J., Dombert, B., Santtila, P., Greenlee, M. W., Osterheider, M., & Mokros, A. (2011). Functional cortical and subcortical abnormalities in pedophilia: A combined study using a choice reaction time task and fMRI. *Journal of Sexual Medicine, 8*, 1660-1674. doi:10.1111/j.1743-6109.2011.02248.x

Price, S. A., & Hanson, R. K. (2007). A modified Stroop task with sexual offenders: Replication of a study. *Journal of Sexual Aggression, 13*, 203-216. doi:10.1080/13552600701785505

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior, 29*, 615-620. doi:10.1007/s10979-005-6832-7

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

Ruch, W. (1999). Die revidierte Fassung des Eysenck Personality Questionnaire und die Konstruktion des deutschen EPQ-R bzw. EPQ-RK [The revised version of the Eysenck Personality Questionnaire and the construction of the German EPQ-R and EPQ-RK]. *Zeitschrift fuer Differentielle und Diagnostische Psychologie, 20*, 1-24. doi:10.1024//0170-1789.20.1.1

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*, 19-30. doi:10.1037/1082-989X.13.1.19

Santtila, P., Antfolk, J., Bäckström, A., Hartwig, M., Sariola, H., Sandnabba, K., & Mokros, A. (under review). *Men's sexual interest in children: One-year incidence and correlates in a population-based sample of Finnish male twins*.

Schaefer, G. A., Mundt, I. A., Feelgood, S., Hupp, E., Neutze, J., Ahlers, C. J., . . . Beier K. M. (2010). Potential and Dunkelfeld offenders: Two neglected target groups for prevention of child sexual abuse. *International Journal of Law and Psychiatry, 33*, 154-163. doi:10.1016/j.ijlp.2010.03.005

Sell, R. L. (1996). The Sell-Assessment of Sexual Orientation: Background and scoring. *Journal of Lesbian, Gay and Bisexual Identity, 1*, 295-310.

Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC: American Psychological Association.

Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The Screening Scale for Pedophilic Interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior, 33*, 455-466. doi:10.1023/B:ASEB.0000037426.55935.9c

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

Singer, B. (1984). Conceptualizing sexual arousal and attraction. *Journal of Sex Research, 20*, 230-240. doi:10.1080/00224498409551222

Smith, P., & Waterman, M. (2004). Processing bias for sexual material: The Emotional Stroop and sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 16*, 163-171. doi:10.1177/107906320401600206

Snowden, R. J., Craig, R. L., & Gray, N. S. (2011). Indirect behavioural measures of cognition among sexual offenders. *Journal of Sex Research, 48*, 192-217. doi:10.1080/00224499.2011.557750

Steffens, M. C., Yundina, E., & Panning, M. (2008). Automatic associations with "erotic" in child sexual offenders: Identifying those in danger of reoffence. *Sexual Offender Treatment, 3*(2). Retrieved from http://www.sexual-offender-treatment.org/2-2008_03.html

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

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99*, 100-117. doi:10.1037/0033-2909.99.1.100

Tatsuoka, M. M. (1971). *Multivariate analysis: Techniques for educational and psychological research.* New York, NY: John Wiley.

Thornton, D., & Laws D. R. (Eds.) (2009). *Cognitive approaches to the assessment of sexual interest in sexual offenders.* Chichester, UK: Wiley.

Wald, A. (1944). On a statistical problem arising in the classification of an individual into one of two groups. *Annals of Mathematical Statistics, 15*, 145-162. doi:10.1214/aoms/1177731280

World Health Organization. (2004). *International statistical classification of diseases and health related problems* (10th rev., Vol. F, 2nd ed.). Geneva, Switzerland: Author.

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:10.1007/s11194-006-9024-1

Worsham, M. (2010). *Creating an expected profile for Affinity 2.5 from a sample of non-pedophilic, exclusively heterosexual, college-age females*. Unpublished doctoral dissertation, Brigham Young University, Provo, UT.

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