Evaluating Parenting Capacity: Validity Problems With the MMPI-2, PAI, CAPI, and Ratings of Child Adjustment

Geoffrey D. Carr, Marlene M. Moretti, and Benjamin J. H. Cue
Simon Fraser University

Practitioners who conduct assessments of parenting capacity for the courts are faced with the challenge of determining the extent to which positive self-presentation by parents distorts test findings. This study examined positive self-presentation bias on commonly used psychological tests in cases referred following removal of children from the home because of abuse or neglect. Substantial positive self-presentation bias was apparent on the measures examined, and parents who presented themselves positively on one test tended to do so on others. Intellectual functioning did not account for these findings. The results demonstrate the pervasive problem of positive self-presentation bias in compromising the validity of test results in this population. Recommendations for conducting clinical assessments with this population are offered, including direction for the use and interpretation of psychological tests.

Keywords: custody evaluations, ethical standards of practice, psychological assessment, test validity

In forensic evaluation, there is likely no area in which emotions run higher than the custody of children (e.g., Otto & Collins, 1995). Parents who are being assessed to aid the courts in determining child custody are, understandably, strongly motivated to present themselves in a positive light, but this can obscure the data on which conclusions must rest. This factor poses significant difficulties for this area of assessment, which clinicians often report to be among the most complex in forensic psychology (e.g., Otto, Edens, & Barcus, 2000).

In recent years, several studies have addressed the problem of positive self-presentation biases in assessing parents in the context of postdivorce child custody. The current study extends previous investigations by examining self-presentation bias in a different population of parents: those being assessed when the question before the courts is whether parental rights should be terminated. These assessments are variously termed *assessments of minimal parenting competence, evaluations for termination of parental rights, child protection evaluations, dependency evaluations,* and as in this report, *parenting capacity assessments* (PCAs). They are requested when the abilities of parents to meet minimal community standards in caring for children are at issue, and typically they relate to a perceived risk of child physical, sexual, and/or emotional abuse or neglect (e.g., Azar, Lauretti, & Loding, 1998; Budd, 2001; Budd & Holdsworth, 1996; Kuehnle, Coulter, & Firestone, 2000).

Researchers and clinicians (e.g., Budd & Holdsworth, 1996; Kuehnle et al., 2000) have lamented the lack of empirical research on PCAs, particularly in light of their impact on the lives of parents and children. There are no published studies on psychological test results in this population, and there are no studies that examine the problem of positive self-presentation across the different types of measures used in assessing parenting capacity. In contrast, in the postdivorce custody literature, there have been reviews of test utilization, presentations of normative data, and examinations of positive self-presentation by parents. It is noteworthy that the reviews of test utilization have consistently indicated that the Minnesota Multiphasic Personality Inventory–2 (MMPI-2) is the most commonly used instrument (e.g., Ackerman & Ackerman, 1997; Hagen & Castagna, 2001; Keilin & Bloom, 1986; Quinnell & Bow, 2001), that normative data on the MMPI-2 have been presented (Bagby, Nicholson, Buis, Radovanovic, & Fidler, 1999; Bathurst, Gottfried, & Gottfreid, 1997; Ollendick & Otto, 1984), and that the issue of defensive responding has been specifically addressed (Bagby et al., 1999; Bathurst et al., 1997; Medoff, 1999; Posthuma & Harper, 1998; Siegel, 1996). These authors have generally reported that MMPI-2 measures of positive self-presentation tend to be slightly or moderately elevated in this population, although the clinical significance of the elevations has been questioned (Medoff, 1999).

GEOFFREY D. CARR obtained his PhD in clinical psychology from McGill University (Montreal, Quebec). He is currently in private practice in Vancouver, British Columbia, and is a clinical associate of Simon Fraser University (Burnaby, British Columbia). His research interests include the evaluation of parenting ability and evaluating the effectiveness of psychotherapy.

MARLENE M. MORETTI received her PhD in psychology from Simon Fraser University and is a professor in the Department of Psychology at Simon Fraser University. Her research focuses on children and families at risk, attachment, program development, and evaluation. Her research is funded through the Canadian Institutes of Health Research, the Social Sciences and Humanities Research Council of Canada, and Heritage Canada.

BENJAMIN J. H. CUE completed his MS in counseling psychology at Simon Fraser University in 2003. He is an intake counselor with a Chicago-based employee assistance and managed behavioral health firm. He continues to pursue a research interest in psychological testing, as well as in the productivity and mental health benefits of employee-assistance programs.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Marlene M. Moretti, Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6. E-mail: moretti@sfu.ca

There are reasons to predict that people assessed in the context of PCAs respond to psychological assessment somewhat differently from those involved in postdivorce custody and access cases. Unlike postdivorce cases, in which the "losing" parent is still likely to gain regular access to the child, in the case of PCAs, the child may be adopted or otherwise unavailable for ongoing contact with the biological parent, further increasing the stakes of the assessment. As noted by others (Azar et al., 1998; Kuehnle et al., 2000), clients being assessed for PCAs are typically of a lower socioeconomic status, lower education, and often lower than average IQ. These factors may affect their responses to assessment questions. For example, the parents in Bathurst et al.'s (1997) postdivorce sample typically had some college education (M = 15.36 years of education), whereas in the present sample only half of the parents had completed Grade 10, and only 15% had any postsecondary education whatsoever.

The Investigation

In the current study, we examined the pattern and level of validity-scale elevations for parents undergoing a PCA, and we discuss how our results compare to those of the postdivorce custody literature. Self-presentation bias was examined on the MMPI-2 (Butcher et al., 2001), the Personality Assessment Inventory (PAI; Morey, 1996), the Child Abuse Potential Inventory (2nd ed.; CAPI; Milner, 1986), and ratings of children's behavior as measured by the Child Behavior Checklist (CBCL; Achenbach, 1991a). Participants included 91 biological mothers and 73 fathers (48 biological fathers and 25 stepfathers) from 93 cases consecutively assessed at Family Court Centre, a government agency for court-ordered PCAs. Referrals originated from judges, lawyers, or social workers. Parents were predominantly of European descent, although a sizable number of participants (9% of mothers and 13% of fathers) were of Aboriginal heritage. Mothers ranged in age from 18 to 53 years (M = 33.1 years, SD = 7.4 years), and fathers ranged from 21 to 60 years (M = 37.4 years, SD = 8.7 years). The majority of participants had not completed their high school education (62% of mothers and 55% of fathers), and most were unemployed and/or on social assistance at the time of the assessment (74% of mothers and 47% of fathers).

In cases in which there was prior concern regarding a parent's ability to understand the meanings of psychological test items, reading ability was typically assessed either formally with a test of reading ability or informally by having the parent read several test items aloud. Intellectual functioning was assessed in 67 mothers and 48 fathers using the Wechsler Adult Intelligence Scale–III (33 mothers and 8 fathers; Psychological Corporation, 1997), the Wechsler Abbreviated Scale of Intelligence (31 mothers and 27 fathers; Psychological Corporation, 1999), or the Shipley Institute of Living Scale (3 mothers and 13 fathers; Zachary, 1994). Combining full-scale estimates across all three of these measures revealed that the mean level of intellectual functioning was in the average range, with fathers (M = 99.6, SD = 12.1) scoring significantly higher than mothers (M = 92.1, SD = 16.7), $F(1, 107) = 6.43$, $p = .01$. Nonetheless, some parents' IQs fell in the mentally retarded range (IQ score of 70 or below for 6 mothers), in the borderline range (71–80 for 13 mothers and 3 fathers), or in the low average range (81–90 for 12 mothers and 3 fathers). These findings are in accord with other published descriptions of parents undergoing PCA (e.g., Azar et al., 1998; Kuehnle et al., 2000), although our results are the first formal assessment and reporting of intellectual and demographic characteristics of this population. In light of the fact that a considerable number of these parents functioned below the average range, we evaluated the relationship between intelligence scores and validity-scale elevations to rule out the possibility that elevations were not simply a function of the inability to understand the test items.

Minnesota Multiphasic Personality Inventory (2nd ed.)

Numerous indices of test validity have been developed for the MMPI-2, including the original Lie (L), Infrequency (F), and Correction (K) scales, which are the focus of the current study. A high score on the L scale indicates a tendency to deny minor faults and complaints; elevations on the K scale suggest a more subtle defensiveness toward the test items; and F scale elevations typically represent increased or exaggerated symptomatology, poor understanding, or careless responding. MMPI-2 profiles revealed frequent elevation of scores on the L scale (M = 62.66, SD = 14.01 for mothers and M = 60.76, SD = 8.69 for fathers), with smaller elevations on the F scale (M = 58.34, SD = 15.13 for mothers and M = 51.43, SD = 9.49 for fathers) and K scale (M = 51.66, SD = 11.78 for mothers and M = 54.60, SD = 11.13 for fathers). Table 1 presents the percentage of MMPI-2 profiles that were elevated on any of the three validity indices for mothers and fathers in the sample. In clinical practice, decisions about MMPI-2

Table 1
Percentages of Elevated MMPI-2 Profiles at t Score Thresholds of 70 and 65

	70 threshold						65 threshold					
	Mothers[a]		Fathers[b]		Total[c]		Mothers[a]		Fathers[b]		Total[c]	
MMPI-2 profile	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*
All valid	50.82	31	71.43	30	59.22	61	36.10	22	45.20	19	39.80	41
L elevated	34.44	21	16.67	7	27.19	28	45.90	28	33.33	14	40.78	42
F elevated	14.75	9	7.14	3	11.65	12	27.90	17	11.90	5	21.40	22
K elevated	6.56	4	4.76	2	5.83	6	18.03	11	23.81	10	20.39	21

Note. Totals of each column row exceed 100% as elevations of each validity scale are considered independently. MMPI-2 = Minnesota Multiphasic Personality Inventory (2nd ed.; Butcher et al., 2001); L = Lie scale; F = Infrequency scale; K = Correction scale.
[a] n = 61. [b] n = 42. [c] n = 103.

protocol validity involve taking a variety of scale scores into consideration, and elevated validity-scale *t* scores of 65 or 70 would not necessarily invalidate a client's test results. For research purposes, however, validity-scale cutoff *t* scores of 65 or 70 on each of the above scales are commonly used (see, e.g., Bagby et al., 1999). To ensure comparability with previous research, the current study includes data on both cutoff thresholds. Of the MMPI-2 profiles, 49% were invalid using the threshold *t* score of 70; elevations were most commonly observed on the L scale, indicating that clients tended to deny faults and problems. With a threshold *t* score of 65, 60% of profiles were invalid. Again, the most common source of invalidity was elevation on the L scale, although approximately one fifth of profiles were elevated for both the F and K scales.

Given the high percentage of profiles that included validity-scale elevations above cutoff thresholds, Table 2 summarizes the extent and significance of validity-scale elevations on the clinical scales. Multivariate analyses of variance (MANOVAs) were completed comparing the clinical scales for profiles with or without elevated validity scales at *t* score thresholds of 65 and 70. Elevations on any of the validity scales resulted in significant effects on the clinical scales at both of the thresholds. L scale elevations at the threshold of 65 resulted in significant effects on four of the clinical scales and two elevations that approached significance. Clients with elevated L scales tended to present themselves as less symptomatic, particularly as less paranoid ($p = .004$ at both thresholds) and less introverted ($p = .02$ and $p = .01$ at the 65 and 70 *t* score thresholds, respectively). Clients with elevated F scale scores predictably obtained significantly higher scores on most of the clinical scales at both thresholds. K scale elevations resulted in clients acknowledging less symptomatology on the hypochondriasis, conversion hysteria, and social introversion scales at both thresholds.

In summary, validity-scale elevations were quite common in this population, with positive self-presentation on L or K compromising the validity of approximately 60% of examinee profiles. Further, this positive self-presentation has a significant suppressive effect on the clinical scale scores. These findings closely replicate our earlier (unpublished) results from a previous cohort of 76 examinees (Moretti, Carr, & Cue, 2002) and extend similar re-

Table 2
Means and Standard Deviations for MMPI-2 Clinical Scale Scores for Profiles at Validity-Scale t Score Thresholds of 65 and 70

	L scale				F scale				K scale			
MMPI-2 clinical scale	<70 (n = 30)	≥70[a] (n = 83)	<65 (n = 47)	≥65[b] (n = 66)	<70 (n = 12)	≥70[c] (n = 102)	<65 (n = 22)	≥65[d] (n = 92)	<70 (n = 7)	≥70[e] (n = 107)	<65 (n = 22)	≥65[f] (n = 92)
Hs												
M	52.19	52.83	52.85	51.66	51.67	59.30**	50.76	58.95**	51.72	61.86*	50.75	59.00**
SD	11.67	10.75	12.70	9.34	10.31	15.26	10.40	12.92	11.26	8.40	11.25	9.27
D												
M	53.38	0.76	54.44	50.28*	51.77	61.70**	51.26	58.41**	52.65	52.43	52.93	51.41
SD	11.70	7.62	12.60	7.02	9.86	11.07	9.96	12.23	11.05	5.38	11.60	6.26
Hy												
M	53.06	53.62	53.39	52.94	53.05	53.50	53.08	53.05	52.33	64.43**	51.22	60.82***
SD	12.02	12.42	12.73	11.22	12.13	11.83	12.20	11.95	11.94	9.03	11.98	9.41
Pd												
M	58.81	54.72	59.27	55.64†	56.51	69.20***	55.80	65.23***	57.24	63.43	57.02	60.14
SD	11.94	8.14	12.44	8.68	10.40	10.72	10.71	10.05	11.29	7.89	11.67	8.66
Mf												
M	52.70	57.55†	52.55	55.91	53.63	57.60	52.93	58.32*	53.82	56.29	54.16	53.18
SD	11.00	10.88	11.42	10.36	11.08	10.17	10.77	11.27	11.19	8.50	13.50	8.60
Pa												
M	57.93	49.83**	58.67	51.89**	53.86	74.40***	53.16	66.09***	55.59	56.71	55.71	55.45
SD	13.00	9.29	12.85	11.10	11.00	12.52	11.16	13.50	13.00	5.99	11.29	10.05
Pt												
M	51.77	49.21	53.17	48.23*	49.77	64.60***	48.87	60.27***	50.75	56.00	50.72	52.55
SD	12.85	8.26	13.74	7.65	9.80	15.04	9.51	15.61	12.02	4.93	12.83	5.39
Sc												
M	54.51	51.14	55.44	51.13†	51.21	77.30***	49.96	68.32***	53.40	55.00	53.55	53.27
SD	14.11	7.04	15.00	7.86	9.41	15.02	9.11	14.95	13.06	4.73	13.93	5.16
Ma												
M	51.87	49.69	51.67	50.81	50.22	61.50**	49.88	56.77**	51.47	47.29	51.54	49.82
SD	10.10	8.74	10.64	8.35	8.85	13.69	8.78	11.68	9.97	3.55	10.07	8.24
Si												
M	50.87	46.21*	51.68	46.85**	48.94	59.50***	48.17	56.95***	50.31	43.14†	51.72	42.14***
SD	9.88	8.77	10.46	8.06	9.35	8.25	9.26	9.89	10.08	4.49	9.96	5.28

Note. MMPI-2 = Minnesota Multiphasic Personality Inventory (2nd ed.); L = Lie; F = Infrequency; K = correction; Hs = Hypochondriasis; D = Depression; Hy = Hysteria; Pd = Psychopathic Deviate; Mf = Masculinity–Femininity; Pa = Paranoia; Pt = Psychasthenia; Sc = Schizophrenia; Ma = Hypomania; Si = Social Introversion.
MANOVAs: [a] $F(10, 102) = 3.01$, $p = .002$. [b] $F(10, 102) = 2.11$, $p = .03$. [c] $F(10, 103) = 7.34$, $p < .001$. [d] $F(10, 103) = 7.39$, $p < .001$. [e] $F(10, 103) = 2.19$, $p = .02$. [f] $F(10, 103) = 5.45$, $p < .001$.
† $p \leq .10$. * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

search in samples of postdivorce child custody cases, although the levels of L and K scale elevations have varied across studies. Studies by Siegel (1996), Bathurst et al. (1997), Posthuma and Harper (1998), and Bagby et al. (1999) have found mean L scale scores to be somewhat elevated (*t* scores of 58, 56, 53, and 52.3, respectively), with elevations on K slightly higher than L in each study (*t* scores of 60, 58.7, 56, and 57.5, respectively). In contrast, our findings show dominant L scale elevations. The discrepancy between our findings and those of previous studies may be accounted for by the higher stakes of termination of parental rights or by the comparatively lower level of education in our sample, because education and socioeconomic status have been found to influence L scale elevations (e.g., Greene, 1991). Our findings on the impact of the validity-scale elevations on clinical scale scores also contrast with Bagby et al.'s (1999) results with a postdivorce custody assessment sample. Although they found that 52% of their sample obtained *t* score elevations of 65 or greater on the L and/or the K scales, they found no effect of these elevations on the clinical scales. The difference between our results and those of Bagby et al. may be due to the fact that underreporting in the current study was typically due to L scale elevations, in contrast to the more frequent K scale elevations found by Bagby and his colleagues. Medoff (1999) noted that the validity-scale elevations among postdivorce child custody clients are statistically significant but not of sufficient proportions to be clinically significant or to suppress clinical scales. This was clearly not the case for the present sample. Many obtained elevations that significantly distorted their scores on clinical scales, such as substantially suppressing scores on the paranoia scale.

In addition to the elevations on the MMPI-2's L and K scales, we found a moderate elevation on the F scale. This is predictable in this population because, in spite of a desire by most parents to present themselves in a positive light and to deny problems, the higher incidence of lower IQ, psychopathology, and cries for help would each contribute to F scale elevations. As presented below, F scale elevations, but not L or K scale elevations, were more frequent in parents with low intellectual functioning. However, this relationship was significant only when using a threshold of 65 to determine validity; thus, factors other than IQ contributed to the observed F scale elevations.

Personality Assessment Inventory

The PAI (Morey, 1996) is a relatively new self-report measure of personality and psychopathology that is being used by a small percentage of psychologists conducting custody and access assessments (Quinnell & Bow, 2001). It includes a number of validity scales, three of which—Infrequency (INF), Positive Impression (PIM), and Negative Impression (NIM)—were examined in the current study. The INF and PIM scales correspond to the MMPI-2's F and L scales, respectively. The NIM scale measures the tendency of respondents to malinger. In the current sample, the measure was completed by 32% of mothers and 30% of fathers, totaling 51 respondents (29 mothers and 22 fathers). The PAI manual (Morey, 1996) specifies that scores of 75 and above on the INF scale, of 66 and above on the PIM scale, and of 92 and above on the NIM scale represent significant elevations, above which interpretation of clinical scales is not recommended. Five out of 29 mothers (17.2%) and 4 out of 22 fathers (18.2%) obtained elevations within the invalid range on the PIM scale. Only 1 mother and no fathers were in the invalid range on the INF scale, and no respondents were in the invalid range on the NIM scale. Although almost one in five PAI profiles were found to be invalid because of positive self-presentation, contrary to predictions, a MANOVA comparing valid and invalid profiles was not significant, $F(11, 39) = 1.375$, *ns*, probably because of the unequal number of respondents in valid ($n = 42$) versus invalid ($n = 9$) profiles.

Child Abuse Potential Inventory (2nd ed.)

The CAPI is a self-report instrument designed to assess the likelihood that a respondent will physically abuse a child in his or her care (Milner, 1986, p. 1). It includes three validity scales: the Lie (L), Random Response (RR), and Inconsistency (IC) scales, which are combined to create three validity indices: Faking Good, Faking Bad, and Random Response. In this sample, 73% of mothers and 64% of fathers completed the CAPI, for a total of 113 respondents (66 mothers and 47 fathers). Of the CAPI profiles, 49% were invalid, with all invalid profiles including invalidation by the Faking Good index. As the CAPI manual indicates, an invalid Faking Good index makes it impossible to interpret normal range Abuse scale scores. A MANOVA confirmed that faking good produced significant distortions on the CAPI scales, $F(8, 93) = 3.44$, $p = .001$. For both mothers and fathers, the mean Abuse scale score for invalid profiles was significantly lower than for valid profiles, $p = .01$. As shown in Table 3, faking good profiles resulted in significantly lower Abuse scale scores and significantly impacted several of the factor scores.

In summary, almost half of the CAPI profiles were considered invalid because of elevations on the Lie scale and resulting Faking Good index scores, and this was associated with significantly lower scores on the CAPI Abuse scale and factor scales. The notable exception was the Rigidity factor scale, which reflects unreasonably high and rigid expectations regarding the behavior and appearance of children, a finding also reported by Milner and Crouch (1997). Even though elevations on this scale are associated with physical abuse, scores were significantly higher for people faking good on the CAPI than for those who did not. Parents who

Table 3
CAPI Abuse and Factor Scale Scores for Valid and Faking Good Profiles

	Valid ($n = 57$)		Faking Good ($n = 54$)	
Scale	*M*	*SD*	*M*	*SD*
Abuse total score**	142.89	100.38	100.65	75.32
Distress**	79.47	75.42	43.20	58.35
Rigidity***	11.12	13.17	20.62	14.82
Unhappiness***	18.88	16.81	9.96	8.61
Problems With Child	9.86	9.09	7.64	7.40
Problems With Family	13.65	11.97	12.34	11.87
Problems With Others	10.16	8.82	8.11	7.49
Ego Strength**	26.35	10.46	31.77	9.67
Loneliness**	5.76	4.34	3.60	3.67

Note. CAPI = Child Abuse Potential Inventory (Milner, 1986).
** $p \leq .01$. *** $p \leq .001$.

physically abuse children tend to have unreasonable expectations for their children, but because they have little insight into this, they do not attempt impression management. This suggests that test items of this type may be particularly valuable for detecting abusive parents even when they attempt to present a positive image.

Child Ratings: Child Behavior Checklist and Teacher Report Form

Parental bias in reporting of child symptomatology has been recognized as a significant validity concern (Webster-Stratton & Lindsay, 1999). In addition to investigating the impact of parental positive self-presentation on how parents portrayed themselves, we also investigated the possibility that this positive presentation bias would be extended to their children, because parents may be motivated to portray their children as more functional and less in need of intervention than may be the case. We accomplished this by comparing child behavior ratings obtained from parents with ratings for the same child obtained from foster parents or teachers. Parents and foster parents completed the CBCL (Achenbach, 1991a), which provides a total problem score in addition to subscale scores reflecting the internalizing and externalizing problems of each child. The Teacher Report Form (Achenbach, 1991b) is a parallel version of the CBCL that is completed by the child's teacher. Unfortunately, the small number of biological ($n = 10$), step ($n = 1$), and foster ($n = 1$) fathers who completed CBCLs precluded their inclusion in these analyses. A MANOVA indicated that biological mothers' ratings of their children's problems were significantly lower than ratings obtained on the same children from their foster mothers ($n = 17$), $F(3, 30) = 5.10$, $p = .006$. A MANOVA revealed a similar trend, with teachers' ratings of children's problems being marginally higher than those of biological mothers ($n = 16$), $F(3, 29) = 2.48$, $p = .08$. A one-way analysis of variance revealed a significant difference ($p = .001$) on the CBCL total scores between mothers ($M = 49.47$, $SD = 10.48$) and foster mothers ($M = 64.18$, $SD = 10.97$). Significant differences were also found on the Internalizing score ($M = 51.88$, $SD = 9.58$ for mothers; $M = 60.06$, $SD = 12.17$ for foster mothers; $p = .04$) and on the Externalizing score ($M = 51.53$, $SD = 10.32$ for mothers; $M = 65.71$, $SD = 12.70$ for foster mothers; $p = .001$). Biological mothers also produced lower ratings than teachers on the total score ($M = 52.06$, $SD = 12.36$ for mothers; $M = 62.06$, $SD = 11.26$ for teachers; $p = .02$), but differences were not significant for the Internalizing score ($M = 52.82$, $SD = 11.29$ for mothers; $M = 57.65$, $SD = 13.12$ for teachers) or the Externalizing score ($M = 54.88$, $SD = 11.31$ for mothers; $M = 63.24$, $SD = 11.98$ for teachers). In summary, these data on child ratings indicate that parents undergoing PCAs are also inclined to present an overly positive picture of their children's functioning. It is noteworthy that a response bias in reporting on the CBCL has previously been reported for parents undergoing court-ordered child custody evaluations (Ash & Guyer, 1991).

Intellectual Functioning and Test Validity

To examine the relationship between intellectual functioning and test validity, we compared the proportion of valid versus invalid profiles among parents with IQ scores of 80 or above to that of parents with IQ scores below 80. No significant association was found between level of intellectual functioning and MMPI-2 L scale validity at a threshold of 70, $\chi^2(1, N = 78) = 0.22$, *ns*, or 65, $\chi^2(1, N = 78) = 2.04$, *ns*. Similarly, intellectual functioning was not associated with K score elevations at the threshold of 70, $\chi^2(1, 78) = 0$, *ns*, or 65, $\chi^2(1, N = 78) = 1.34$, *ns*. Predictably, some evidence was found for an association between elevations on the F scale and intellectual functioning. A trend emerged showing that F scale scores were more frequently elevated at or beyond the threshold of 70 for parents with lower levels of intellectual functioning, as compared with parents at higher levels of intellectual functioning, $\chi^2(1, N = 78) = 2.79$, $p = .10$. This trend was significant when the threshold was dropped to 65, $\chi^2(1, N = 78) = 3.71$, $p = .05$.

For validity indices on other tests, we found no evidence that low intellectual functioning was associated with faking good on the CAPI, $\chi^2(1, N = 69) = 0.24$, *ns,* or with positive impression management on the PAI, $\chi^2(1, N = 78) = 2.49$, $p = .11$. In fact, in the case of the PAI, the relationship was reversed, with invalid PIM scores more frequently occurring in higher functioning parents (32%) than in low functioning parents (0%). In summary, no evidence was found to suggest that low intellectual functioning accounts for the pattern of positive self-presentation consistently found across multiple tests in this study.

Relation of Validity Measures Across Instruments

Given the consistent indications of positive self-presentation across different measures, we endeavored to determine if the different measures of this response bias were due to a common factor. As presented in Table 4, our first examination involved correlations across the validity indices. For comparison with measures of validity from the other psychological tests, the child-ratings data were calculated as difference scores, with the CBCL score for the biological mother being subtracted from the rating of the child by the foster mother or teacher. The MMPI-2's L scale was significantly correlated with both the CAPI's L scale, $r(80) = .60$, $p > .001$, and the PAI's PIM scale, $r(28) = .66$, $p < .001$. Similarly, the MMPI-2's K scale was also significantly correlated with both the CAPI's L scale, $r(80) = .37$, $p = .001$, and the PAI's PIM scale, $r(28) = .73$, $p < .001$. Not surprisingly, the CAPI's L scale and the PAI's PIM scale exhibited a significant correlation as well, $r(40) = .47$, $p < .001$. Elevations on the MMPI-2's L scale were marginally correlated with the child-ratings difference scores for both the total scores, $r(14) = .49$, $p = .07$, and the Externalizing scores, $r(14) = .50$, $p = .07$. Most of the other correlations between the CBCL/Teacher Report Form difference scores and the validity indicators of other measures did not reach statistical significance because of limited power.

A second examination of the relationships among the different validity measures assessed cross-test consistency: whether invalidity on one measure was predictive of invalidity on other measures. Results show that elevations of the MMPI-2's L scale at or above the threshold of a *t* score of 65 were significantly associated with faking good on the CAPI, $\chi^2(1, N = 75) = 22.20$, $p = 0$. Similarly, elevation of the MMPI-2's L scale was associated with invalidity on the PAI as measured by the PIM validity index, although this failed to reach significance, $\chi^2(1, N = 26) = 2.73$,

Table 4
Intercorrelations Between Validity Indices on the MMPI-2, PAI, CAPI, and CBCL/TRF Difference Scores

Measure	1	2	3	4	5	6	7	8	9	10	11
1. MMPI-2 L	—	−.26**	.53***	.66***	−.28	.37*	.60***	.13	.29	.50†	.49†
N		102	102	28	28	28	80	79	14	14	14
2. MMPI-2 F		—	−.50***	−.50*	.34†	.39*	−.23†	.02	−.43	−.59*	−.60*
N			102	28	28	28	80	79	14	14	14
3. MMPI-2 K			—	.73***	−.72***	−.01	.37**	.01	.27	.33	.38
N				28	28	28	80	79	13	13	13
4. PAI PIM				—	−.34*	.18	.47**	.12	.31	.04	.24
N					47	47	40	40	7	7	7
5. PAI NIM					—	−.06	−.08	−.27†	−.47	.02	−.27
N						50	50	43	7	7	7
6. PAI INF						—	.09	.29†	.04	−.20	−.02
N							43	43	7	7	7
7. CAPI L							—	.25**	−.23	−.17	−.08
N								111	13	13	13
8. CAPI RR								—	−.54†	−.42	−.35
N									13	13	13
9. CBCL/TRF Internalizing difference score									—	.80***	.88***
N										22	22
10. CBCL/TRF Externalizing difference score										—	.88***
N											22
11. CBCL/TRF total scale difference score											—

Note. MMPI-2 = Minnesota Multiphasic Personality Inventory–2 (Butcher et al., 2001); L = Lie scale; F = Infrequency scale; K = Correction scale; PAI = Personality Assessment Inventory (Morey, 1996); PIM = Positive Impression scale; NIM = Negative Impression scale; INF = Infrequency scale; CAPI = Child Abuse Potential Inventory (Milner, 1986); RR = Random Response scale; CBCL = Child Behavior Checklist (Achenbach, 1991a); TRF = Teacher Report Form (Achenbach, 1991b).
† $p \leq .10$. * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

$p = .10$. PIM invalid PAI profiles were not, however, significantly associated with faking good on the CAPI, $\chi^2(1, N = 40) = 1.95$, *ns*. Elevation on the MMPI-2's L scale was therefore a robust predictor of invalidity on other psychological tests, notably the CAPI and PAI.

In summary, the different measures of positive self-presentation are all positively related to each other, although they are only somewhat overlapping and can result in different conclusions about test validity. Invalidity on the MMPI-2 and invalidity on the CAPI were significantly correlated, and the threshold for determining validity on the two measures resulted in close correspondence in the classification of invalid profiles. Seventy-eight percent of examinees who produced valid MMPI-2 profiles were also found to produce valid CAPI profiles, and 77% of examinees who produced invalid MMPI-2 profiles also produced invalid CAPI profiles. In contrast to the MMPI-2 and CAPI, which identified half or more of the test protocols as invalid, the PAI identified less than 20% of respondents as having invalid profiles. Although both the MMPI-2's L scale and the CAPI Faking Good index correlated significantly with the PAI's PIM scale, classification of protocols as invalid on the MMPI-2 or the CAPI did not correspond with classification of invalidity on the PAI. Sixty-four percent of examinees who produced an invalid MMPI-2 protocol were found to produce a valid PAI protocol. Similarly, 58% of examinees who produced an invalid CAPI protocol were found to produce a valid PAI protocol. The PAI's PIM scale therefore appears to be a fairly conservative criterion for assessing response bias, and it will tend to miss cases of positive self-presentation identified by other measures.

The relationships among the infrequency-type scales were also noteworthy. The MMPI-2's F scale was shown to be significantly correlated with the PAI's INF scale, $r(28) = .39$, $p < .05$. Surprisingly, the CAPI's RR scale was not significantly correlated with the MMPI-2's F scale and was only marginally correlated with the PAI's INF scale, $r(43) = .29$, $p < .10$. Significant correlations in the predicted direction were found between the MMPI-2's F scale and the CBCL/Teacher Report Form Externalizing difference score, $r(14) = –.59$; $p = .03$, and total difference score, $r(14) = –.60$; $p = .02$, likely reflecting a general tendency among high scorers to acknowledge problems.

Conclusions and Recommendations for Practice

These data represent the first comprehensive presentation of psychological test findings for parents undergoing PCAs. It is clear that positive self-presentation is a significant factor affecting responses to psychological tests in this population, apparently affecting test results more substantially for this group than for parents undergoing postdivorce custody assessments. Collectively, the results demonstrate that this positive bias is pervasive, as evidenced on each measure we examined, including measures of personality, parenting attitudes, and ratings of children. The results also demonstrate that this positive response bias usually had a significant impact on the tests' clinical scales, resulting in positively distorted and invalid profiles. Finally, elevations of validity indices across psychological tests were correlated, suggesting that validity indices on the different measures are tapping a similar construct.

The finding that people participating in PCAs often present themselves in an improbably positive light is open to two basic interpretations. The first is that this reflects the demands of the situation, in which regaining custody of children creates a power-

ful motive to present oneself positively. The second is that this finding reflects trait characteristics of this population. Our data do not provide information to evaluate these two possibilities, but data from Milner and Crouch (1997) suggest that characteristics of the population may play a role. In their study, even when at-risk parents were given instructions to complete the CAPI honestly and were given a promise of anonymity, 33% were still identified by the Faking Good index, as compared with only 8% of the general population sample. However, on the basis of clinical experience and the results reported in the related literature on postdivorce child custody assessment, we believe that the influence of the situational factor is substantial.

Although most professionals engaged in assessing parenting capacity would certainly acknowledge that parents are motivated to present themselves positively, it is difficult to know how well they take this into account in arriving at conclusions and recommendations. It would obviously be detrimental to the welfare of children if parents' success in receiving custody were heavily influenced by successfully presenting themselves in an unrealistically positive light. There are, however, data that point in this direction. Otto and Collins (1995) reported that in the context of custody and access decisions, parents receiving custody obtained significantly higher elevations on the MMPI-2's K scale (subtle defensiveness) than parents who did not receive custody. Faking good in child custody assessments may therefore be an effective strategy for gaining custody, one that clinicians are compelled to be aware of and about which the courts need to be informed. This is particularly important considering that there is evidence that parents who engage in negative behaviors, such as alienating children from their expartners, are also prone to producing greater elevations on the MMPI-2's K scale (Siegel & Langford, 1998).

The finding of positive self-presentation has several implications for practice in this field. Of greatest importance are the implications of these findings for guiding the interpretation of psychological test protocols with elevated validity scales. It could be argued that the findings of the present study simply present new validity-scale norms for this population and that therefore apparent faking good profiles should be accepted as normal, allowing a normal interpretation of the clinical scales. We discourage this interpretation of the results. Although the demands of the assessment situation pull for positive self-presentation, the same tendencies will result in the clinical scales being of limited usefulness. If validity-scale elevations invalidate the test results for other groups, we recommend that the same apply here. However, the validity-scale elevations can reasonably be understood as resulting from the demands of the test situation, and conclusions about client traits based on these elevations may be ill founded.

Given that parents undergoing PCAs are prone to positive self-presentation, the main question is how to deal with it. Some clinicians have concluded that the problem with validity-scale elevations can be dealt with by simply not administering psychological tests. We recommend against this conclusion, which we regard as akin to disconnecting an annoying smoke detector. Although there are limitations inherent in the use of psychological tests for assessing parenting capacity, we believe the pitfalls of conducting assessments without psychological tests that include validity indices are obvious. In contrast to clinicians' beliefs about their abilities, their capacity to detect deceit is notoriously poor (e.g., Ekman & O'Sullivan, 1991). In the absence of more sophisticated measures of response bias, clinicians will be prone to errors in assessing the validity of information provided by these clients. We therefore recommend that the psychological tests with validity scales be used in PCAs. The problem of positive self-presentation might be avoided by warning clients prior to testing that validity scales are designed to detect positive self-presentation and that biased responding is not in clients' interests. The drawback of this approach is that it deviates from the standardized administration procedures, and although it may be justified, the effects on the test results are unknown. The approach that we recommend is to inform the client after the invalid test results have been obtained about the results and offer them the opportunity to complete the testing again. Although not always possible for practical reasons, the combined results from the two test administrations provide information about response biases, the capacity to alter such biases following feedback, and possibly valid test results on the second testing. This approach is recommended by the CAPI test author (see Milner, 1986) and by the MMPI-2 test authors (see Graham, 1989, and Butcher, Morfitt, Rouse, & Holden, 1997). Although Butcher et al. (1997) have reported that airline pilot job applicants given this opportunity produce valid results at the second testing, Geoffrey D. Carr has also found striking examples of parents who produced test profiles with virtually identical validity-scale scores on the second administration.

The present results also indicate that psychological test results from instruments that do not have validity scales are of questionable value in PCAs. We recommend that the use of such tests be avoided and that, when their use is clinically compelling, the validity of the results be assessed in light of the validity-scale results from other measures, such as the MMPI-2 and the CAPI. Support for this approach comes from Field and her colleagues (see, e.g., Field, 1995), who found that depressed mothers who denied depressive symptoms on the Beck Depression Scale–II (Beck, Steer, & Brown, 1996) also faked good on the MMPI-2. These faking good mothers were actually less responsive to their children than depressed mothers who acknowledged their symptoms, but without the MMPI-2 validity-scale data, low Beck scores could be misinterpreted as indicating a true lack of depression and low risk to the child.

The pitfalls of using psychological tests without strong validity measures are illustrated with the Adult–Adolescent Parenting Inventory (AAPI; Bavolek, 1984), a test commonly used in the assessment of parenting capacity. In our previous, unpublished results, we found that our client population obtained scores on the AAPI that were superior to those of the normative sample of nonabusive parents (Moretti, Carr, & Schoular, 1999). Milner and Crouch (1997) have similarly demonstrated that at-risk parents are capable of faking good on the AAPI, obtaining scores at or better than the normative sample. Further, they found that a measure of positive self-presentation on another commonly used parenting measure (the Defensive Responding Scale of the Parenting Stress Index; Abidin, 1995) was quite poor at detecting positive self-presentation in at-risk parents. Collectively, the present findings and those of previous research illustrate that failing to use tests with effective validity measures leaves clinicians highly vulnerable to drawing inaccurate conclusions in PCAs.

In light of our findings and related research, we recommend the following for current practice in conducting PCAs. Considering (a) that self-presentation bias is a significant problem with parents undergoing PCAs, (b) that there is extensive research on the MMPI-2 validity scales, (c) that our research suggests the MMPI-

2's L scale is the most robust measure of positive self-presentation in this population, (d) that additional MMPI-2 validity scales (e.g., the Superlative scale) have recently been included in the standard scoring, and (e) that valid validity-scale data are crucial to interpreting other data gathered during PCAs, we recommend that the MMPI-2 be used routinely in these assessments. We recommend that the validity-scale results be recognized as being continuous and therefore as providing information about the degree of the positive self-presentation bias, even in cases with subcutoff elevations. We believe that results obtained on the validity scales of the MMPI-2 and other tests can be used to provide crucial information regarding an individual's approach to the assessment, which can reasonably be assumed to apply to other tests and to some extent to the interview situation as well (e.g., Otto & Collins, 1995, Posthuma & Harper, 1998). At this point, we believe that the CAPI, which was specifically designed to detect parents at risk to physically abuse children, has the best validity scales to support its use in this population. Finally, the likelihood of positive self-presentation in this population highlights the importance of clinicians gathering information about the parents and children from multiple sources (i.e., different types of psychological testing, interviews, observations, third-party information) in order to maximize the likelihood of drawing valid conclusions.

Regarding future practice in this area, ideally psychologists would develop and use assessment techniques that are less susceptible to response bias. Arguments have been made that projective techniques, such as the Rorschach Inkblot Method, which are much less susceptible to such bias, are useful in complementing objective personality tests in such situations (e.g., Weiner, 1999). Indeed, the Rorschach and other projective techniques are often used in postdivorce child custody assessments (Ackerman & Ackerman, 1997; Hagen & Castagna, 2001; Keilin & Bloom, 1986; Quinnell & Bow, 2001). These authors also reported the common use of standardized intelligence tests in custody and access cases, and along with the clinical information provided, these tests have the benefit of being immune from the positive self-presentation bias because the demand characteristics of intelligence tests are for people to do their best. We believe that the field would advance tremendously with the development of standardized instruments for assessing parenting capacity that, analogous to intelligence tests, ask parents to do their best on tasks that are directly related to parenting. We are encouraged by recent efforts to develop standardized scoring and normative data for the Marschak Interaction Method (see, e.g., Jernberg, Booth, Koller, & Allert, 1991), an in vivo clinical tool that taps different parenting skills and is well suited to these assessments.

References

Abidin, R. R. (1995). *Parenting Stress Index.* Odessa, FL: Psychological Assessment Resources.

Achenbach, T. M. (1991a). *Manual for the Child's Behavior Checklist/4–18 and 1991 profile.* Burlington, VT: University of Vermont.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 profile.* Burlington, VT: University of Vermont.

Ackerman, M. J., & Ackerman, M. C. (1997). Custody evaluation practices: A survey of experienced professionals (revisited). *Professional Psychology: Research and Practice, 28,* 137–145.

Ash, P., & Guyer, M. J. (1991). Biased reporting by parents undergoing child custody evaluations. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 835–838.

Azar, S. T., Lauretti, A. F., & Loding, B. V. (1998). The evaluation of parental fitness in termination of parental rights cases: A functional–contextual perspective. *Clinical Child and Family Psychology Review, 1,* 77–100.

Bagby, R. M., Nicholson, R. A., Buis, T., Radovanovic, H., & Fidler, B. J. (1999). Defensive responding on the MMPI-2 in family custody and access evaluations. *Psychological Assessment, 11,* 24–28.

Bathurst, K., Gottfried, A. W., & Gottfried, A. E. (1997). Normative data for the MMPI-2 in child custody litigation. *Psychological Assessment, 9,* 205–211.

Bavolek, S. J. (1984). *Handbook for the Adult–Adolescent Parenting Inventory (AAPI)* Schaumburg, IL: Family Development Associates.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory–II.* San Antonio, TX: Harcourt Brace.

Budd, K. S. (2001). Assessing parenting competence in child protection cases: A clinical practice model. *Clinical Child and Family Psychology Review, 1,* 1–18.

Budd, K. S., & Holdsworth, M. J. (1996). Issues in clinical assessment of minimal parenting capacity. *Journal of Clinical Child Psychology, 25,* 2–14.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory–2) manual for administration, scoring, and interpretation* (rev. ed.). Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Morfitt, R. C., Rouse, S. V., & Holden, R. R. (1997). Reducing MMPI-2 defensiveness: The effect of specialized instructions on retest validity in a job applicants sample. *Journal of Personality Assessment, 68,* 385–401.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46,* 913–920.

Field, T. (1995). Psychologically depressed parents. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 4. Applied and practical parenting* (pp. 85–99). Mahwah, NJ: Erlbaum.

Graham, J. R. (1989). Increasing the likelihood of obtaining valid MMPI-2 profiles. In J. N. Butcher & J. R. Graham (Eds.), *Topics in MMPI-2 interpretation* (pp. 26–27). Minneapolis, MN: MMPI-2 Workshops and Symposia, Department of Psychology, University of Minnesota.

Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual.* Boston, MA: Allyn & Bacon.

Hagen, M. A., & Castagna, M. (2001). The real numbers: Psychological testing in custody evaluations. *Professional Psychology: Research and Practice, 32,* 269–271.

Jernberg, A., Booth, P., Koller, T., & Allert, A. (1991). *Manual for the administration and the clinical interpretation of the Marschak Interaction Method (MIM) Preschool and School Age.* Chicago: Therapy Institute.

Keilin, W. G., & Bloom, L. J. (1986). Child custody evaluation practices: A survey of experienced professionals. *Professional Psychology: Research and Practice, 17,* 338–346.

Kuehnle, K., Coulter, M., & Firestone, G. (2000). Child protection evaluations: The forensic stepchild. *Family and Conciliation Courts Review, 38,* 368–391.

Medoff, D. (1999). MMPI-2 validity scales in child custody evaluations: Clinical vs. statistical significance. *Behavioral Sciences and the Law, 17,* 409–411.

Milner, J. S. (1986). *The Child Abuse Potential Inventory: Manual* (2nd ed.). Webster, NC: Psytec.

Milner, J. S., & Crouch, J. L. (1997). Impact and detection of response distortions on parenting measures used to assess risk for child physical abuse. *Journal of Personality Assessment, 69,* 633–650.

Moretti, M., Carr, G. D., & Cue, B. J. H. (2002). [Family Court Centre: Evaluation of client problems and psychological profiles, clinical recommendations, and perceived program efficacy]. Unpublished raw data.

Moretti, M., Carr, G. D., & Schoular, D. (1999). [A review of cases assessed at Family Court Centre: Demographic and psychological profiles, referral concerns, and clinical recommendations]. Unpublished raw data.

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory.* Odessa, FL: Psychological Assessment Resources.

Ollendick, D. G., & Otto, B. J. (1984). MMPI characteristics of parents referred for child-custody studies. *Journal of Psychology, 117,* 227–232.

Otto, R. K., & Collins, R. P. (1995). Use of the MMPI-2/MMPI-A in child custody evaluations. In Y. S. Ben-Porath, J. R. Graham, G. C. N. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.), *Forensic applications of the MMPI-2* (pp. 222–252). Thousand Oaks, CA: Sage.

Otto, R. K., Edens, J. F., & Barcus, E. H. (2000). The use of psychological testing in child custody evaluations. *Family and Conciliation Courts Review, 38,* 312–340.

Posthuma, A. B., & Harper, J. F. (1998). Comparison of MMPI-2 responses of child custody and personal injury litigants. *Professional Psychology: Research and Practice, 29,* 437–443.

Psychological Corporation. (1997). *Wechsler Adult Intelligence Scale—Third Edition manual.* San Antonio, TX: Harcourt Brace.

Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence manual.* San Antonio, TX: Harcourt Brace.

Quinnell, F. A., & Bow, J. N. (2001). Psychological tests used in child custody evaluations. *Behavioral Sciences and the Law, 19,* 491–501.

Siegel, J. C. (1996). Traditional MMPI-2 validity indicators and initial presentation in custody evaluations. *American Journal of Forensic Psychology, 14,* 55–62.

Siegel, J. C., & Langford, J. S. (1998). MMPI-2 validity scales and suspected parental alienation syndrome. *American Journal of Forensic Psychology, 16,* 5–14.

Webster-Stratton, C., & Lindsay, D. W. (1999). Social competence and conduct problems in young children: Issues of assessment. *Journal of Child Clinical Psychology, 28,* 25–43.

Weiner, I. B. (1999). What the Rorschach can do for you: Incremental validity in clinical applications. *Assessment, 6,* 327–340.

Zachary, R. A. (1994). *Shipley Institute of Living Scale revised manual.* Los Angeles: Western Psychological Services.