

Ordinary Physical Punishment: Is It Harmful? Comment on Gershoff (2002)

Diana Baumrind
University of California, Berkeley

Robert E. Larzelere
Nebraska Medical Center

Philip A. Cowan
University of California, Berkeley

E. T. Gershoff (2002) reviewed processes that might mediate and contexts that might moderate the associations between corporal punishment (CP) and child behaviors and provided an account of the methodological weaknesses of the research reviewed in her meta-analyses. In this examination of Gershoff, the authors argue that the biases and confounds in the meta-analyses further limit any causal inferences that can be drawn concerning the detrimental “effects” of CP on associated child behaviors. The authors suggest that undesirable child outcomes are associated with CP because the construct marks inept harsh parenting and conclude that although the harmful effects of physical abuse and other extreme punishments are clear, a blanket injunction against spanking is not justified by the evidence presented by Gershoff.

In the first section of her comprehensive article, Gershoff (2002) used meta-analyses of 88 studies to provide what she offered as a conclusive answer to the question of whether corporal punishment (CP) is significantly associated with 11 frequently identified aspects of children’s behavior. She found that except for an association with immediate compliance to parental demands, CP is associated with negative or undesirable behaviors (e.g., aggression, lower levels of moral internalization and mental health). The second and third sections of the Gershoff review were well differentiated and theoretically valuable. In the second section, she endeavored to explain how CP might cause the child outcomes on which she reported and she provided a sophisticated conceptual analysis of proposed mediators as explanatory factors and of possible social and family context moderators of the links between

CP and these outcomes. Her third section documented the critical methodological inadequacies of the primary studies that in our view call into question the validity and significance of any conclusion that can be drawn from the meta-analyses presented in the first section.

Measurement Issues: Corporal Punishment, Spanking, and Abuse

Disputes about the potential impact of corporal punishment are occurring in the social context of a vigorous debate about the impact of spanking. There is an international movement by lay organizations such as EPOCH-Worldwide (EPOCH-USA, 2000) to criminalize corporal punishment or to discourage its use (American Academy of Pediatrics, 1998; Bitensky, 1997; Hyman, 1996; Straus, 1994). The advocates of a blanket injunction against spanking (e.g., Garbarino, 1996; Holden, Thompson, Zambarano, & Marshall, 1997; Straus, 1994, 1999) have cited associations between physical punishment and detrimental child outcomes, such as those presented in Gershoff’s (2002) meta-analyses, to support their position. Such direct application of correlational research to social policy decisions presupposes a causal link between physical punishment and the detrimental child behaviors with which it is associated. Before we deal with the issues of causal inference, it is necessary to examine what Gershoff meant by *corporal punishment* to be clear about how it is being defined and measured.

There is a sense in which participants in the current debate about the effects of corporal punishment are talking past each other. One issue to which some attention has been given is the need to make a distinction between legal abuse and corporal punishment. A second issue, which has not been given the attention it deserves, is the distinction between harsh and punitive but not legally designated abusive punishment and the more moderate application of

Editor’s Note. W. Andrew Collins served as the action editor for this article.—NE

Diana Baumrind and Philip A. Cowan, Institute of Human Development, University of California, Berkeley; Robert E. Larzelere, Department of Psychology, Nebraska Medical Center, Omaha, Nebraska.

We wish to acknowledge Elizabeth Thompson Gershoff’s generous assistance, limited only by her prior heavy commitments, to help us understand the rules she used to select variables and to perform the meta-analyses.

Correspondence concerning the statistical analyses included in this article and requests for full documentation of these analyses should be addressed to Robert E. Larzelere, Department of Psychology, Munroe-Meyer Institute, 985450 Nebraska Medical Center, Omaha, Nebraska 68198-5450. E-mail: rlarzelere@unmc.edu. Other correspondence concerning this article should be addressed to Diana Baumrind or Philip A. Cowan, Institute of Human Development, University of California, Berkeley, 1203 Tolman Hall #1690, Berkeley, California 94720-1690.

normative spanking within the context of a generally supportive parent-child relationship. Although Gershoff cited Baumrind and Larzelere as concluding "that corporal punishment is both effective and desirable" (Gershoff, 2002, p. 539), it would be more correct to say they concluded only that a blanket injunction against disciplinary spanking is not warranted by the data (e.g., Baumrind, 1996a).

An important scientific consensus conference on corporal punishment defined *spanking* as that subset of the broader category of corporal punishment that is "a) physically non-injurious; b) intended to modify behavior; and c) administered with an opened hand to the extremities or buttocks" (Friedman & Schonberg, 1996, p. 853). Conferees, including Baumrind and Larzelere, agreed that the important debate was not about overly severe forms of corporal punishment but only about parental spanking, which inflicts a minor, temporary level of physical pain, if that. Conferees agreed that abusive corporal punishment should be criminalized and corporal punishment that exceeds mild to moderate spanking or is used with children younger than 18 months or after puberty should be strongly discouraged. Conferees also agreed that a mild spank as a "back-up" to other disciplinary tactics (e.g., reasoning, time-out) could increase the effectiveness of these alternative disciplinary tactics in preschoolers with behavior problems. There was agreement that by itself spanking cannot accomplish the longer term objective of parents—to promote children's competence, moral character, and mental health (see Baumrind 1996a, 1996b, for a further discussion of this point)—and that frequent or severe punishment of any kind signals the presence of a problem in the family dynamics. The issue that was and remains controversial is whether mild to moderate disciplinary spanking (in the years between 18 months and puberty) has been shown to be harmful.

Gershoff's (2002) meta-analyses, however, do not address that issue because her conceptual and operational definitions of *corporal punishment* included punishment that was often too severe and was thus a proxy for the harsh, punitive discipline that is acknowledged by all experts to be detrimental to children's well-being and ethically unacceptable. In the fifth moderator variable in the note to her Table 3 (indexing how researchers operationalized corporal punishment), Gershoff included "severity" as Category 2 and "frequency and severity" as Category 3, indicating that severity (as well as frequency) contributed to the CP measure in those studies so categorized. But her other categories, such as "frequency" (Category 1), did not exclude overly severe forms of CP when the specific question asked of the respondent was about their frequency (E. T. Gershoff, personal communication, October 12, 2001). This meant that slapping in the face or beating with a stick (Engfer & Schneewind, 1982; Eron, 1982; Lefkowitz, Walder, & Eron, 1963; Mahoney, Donnelly, Lewis, & Maynard, 2000) or hitting, pushing, grabbing, or shoving (Simons, Johnson, & Conger, 1994) were not coded as severe when the studies asked about their frequency.

Straus (1994) claimed that the physical abuse or Severe Violence items in the Conflict Tactics Scale can be used "to partial out physical abuse in a statistical analysis or to remove abused children from the sample in order to avoid confounding corporal punishment with physical abuse" (p. 202). Although Straus's own studies rarely did that, he clearly agreed with us that it is proper to do so. However, Gershoff (2002) indicated in her Table 1 that 16%

of the 88 studies in her review operationalized CP as "hit with an object," which Straus (1994) included in his index of abuse or Severe Violence on the Conflicts Tactics Scale. Thus, some of Gershoff's CP measures included behaviors that are part of Straus's (1990, 1994) abuse construct.

To determine the influence of overly severe CP on the results of Gershoff's (2002) meta-analyses, Larzelere coded the severity of CP for 64 primary studies, including 52 of the 54 studies in the aggression composite (i.e., studies measuring aggression or anti-social behavior in children or aggression, criminal behavior, or physical abuse in adults). Baumrind also coded 41% of the studies for reliability, which yielded good interrater agreement ($\kappa = .91$) for severe versus not overly severe CP. Almost two thirds (65.4%) of the 52 aggression composite studies used overly severe CP. Variables were categorized as overly severe if there was direct evidence of severity, such as a CP measure based only on hitting with a belt or stick (30.8% of the 52 studies in Gershoff's aggression composite; e.g., Holmes & Robins, 1988); if severity was part of the definition of the highest CP score (28.8%; e.g., Sears, Maccoby, & Levin, 1957); or if severity was inferred from the label, such as *punitive discipline* (5.8%; e.g., Kandel, 1990). That left only 34.6% of the 52 effect sizes in the aggression composite for which the CP measure was not influenced directly by overly severe CP.

This problem of overly severe (and overly broad) CP can best be illustrated with quotes from Gershoff's primary studies. The following descriptions were used to describe at least part of their definition of CP: "slapped on face, head, and ears" and "shook" (Mahoney et al., 2000, pp. 271–272); "severity of punishment for aggression to parents" (Yarrow, Campbell, & Burton, 1968 [Cited as Radke-Yarrow, Campbell, & Burton, 1968, in Gershoff, 2002], pp. 72); "throwing something at the child" and "severe, strict, often physical" as contrasted with "nonrestrictive, mostly positive guidance," (Deater-Deckard, Dodge, Bates, & Pettit, 1996, p. 1067); usage of "switch, belt, razor strap, paddle, buggy whip, boxing ears" (Caesar, 1988, p. 55); and "hit with belt, stick" (Holmes & Robins, 1988, p. 31). In all of these five studies, an effect size could have been based on a CP measure that did not include such overly severe components of CP. For example, Mahoney et al. (2000) presented separate data for six tactics included in the revised Conflict Tactics Scale's CP measure. The corresponding effect sizes (d) ranged from 0.08 for "spanked bottom with bare hand" to 0.58 for "slapped on face, head, and ears."

In other studies, effect sizes could only be based on CP measures contaminated by overly severe components. Examples include "slaps in the face" and "beating with a stick, a belt, etc." (Engfer & Schneewind, 1982, p. 133); "slap him in the face" and "wash out his mouth with soap" (Lefkowitz et al., 1963, p. 161); "How often were you beaten by your mother (father)?" (Riggs & O'Leary, 1996, p. 526); "kicked, bit, or hit you with a fist," causing "bruises or cuts," and six more violent items (Muller, 1996, pp. 477–478); "rough handling, shaking" (Gordon & Smith, 1965, p. 655); "mom (dad) was a violent or physically abusive person" (Simons, Johnson, Beaman, & Conger, 1993, p. 716); and "severe punishment, parents very angry or hostile, beatings, . . . 'Punished him so he wouldn't forget it'" (Sears et al., 1957, p. 248, which was the CP measure in Levin & Sears, 1956).

In at least one other primary study in the meta-analyses, a large majority of those who were physically punished were also physi-

cally abused. In Lester's (1991) study of inmate records, 49% of those who had attempted suicide had been physically punished by their fathers, but almost as many (44%) had been physically abused. Therefore, it would appear that only 5% of the inmates were physically punished without being abused.

We thus conclude that the meta-analyses included studies of overly severe CP to a far greater extent than a casual reading of the selection criteria would indicate (e.g., Gershoff, 2002, Table 1). To address the current debate about mild to moderate spanking, Gershoff would have had to exclude CP that is excessive, extreme, or cruel. But Gershoff only excluded CP that "knowingly would cause severe injury to the child" (Gershoff, 2002, p. 543), although (according to Gershoff) corporal punishment constitutes abuse in definitions of 22 states not merely when it is intended to result in "significant" physical injury but when it is "excessive," "cruel," "extreme," or "severe."

Beyond the question of measurement of CP, there is an important question of meaning. In her review, Gershoff stated that many studies ask parents about their use of CP but do not define what the researchers mean and do not ascertain what the participants mean. The question is whether it makes sense to aggregate all these studies in a meta-analysis when the definition of the key variable measuring CP is so varied and ambiguous.

Threats to the Validity of the Meta-Analyses

With the exception of the links between CP and immediate compliance or physical abuse, Gershoff (2002) disavowed causal implications of the meta-analyses she presented. However, it is primarily within a causal context that meta-analysis is judged to be superior to a conventional review for theory development (Cook et al., 1992; Greenland, 1998; Miller & Pollock, 1994) and application to social policy (Mann, 1994). The very term *effect size* implies a causal relationship. If, for example, an equally plausible argument can be made that child aggression is a contributing cause of CP as can be made that CP is a contributing cause of child aggression, then it is arbitrary to treat CP as though it is the independent variable and child aggression as the dependent variable, and certainly without first establishing temporal order. Yet antispanking advocates frequently draw causal implications from the associations between CP and child behaviors. For example, on the basis of similar associations, Straus (1994) projected that a reduction in CP would "result in fewer people who are alienated, depressed or suicidal, and in fewer violent marriages" (p. 219) and "include lower crime rates . . . and less money spent on controlling or treating crime and mental illness" (p. 188).

Methodologists in epidemiology and behavioral science have emphasized the importance of empirically establishing the causal status of a correlate (Campbell & Stanley, 1966; Cook & Campbell, 1979; Kazdin, Kraemer, Kessler, Kupfer, & Offord, 1997; Kraemer et al., 1997; Mrazek & Haggerty, 1994) by first showing that it is a risk factor and then that it is a causal risk factor. A *risk factor* (causal or not) must be shown to precede the outcome and to be associated with an increase in the outcome compared with its population base rate. An experimental or prospective longitudinal design is a prerequisite for establishing a temporal sequence by showing that the outcome has not occurred prior to, or simultaneously with, the presumed risk factor. Although, as Greenland (1998) stated, "No meta-analysis can compensate for the inherent

limits of non-experimental data for making inferences about causal effects" (p. 644), nonexperimental studies differ greatly in their research quality and thus in their relevance to a causal argument.

We next consider threats to the validity first of the primary studies in Gershoff's (2002) analyses and then of the synthesis itself that may have systematically inflated the effect size estimates or compromised the generalizability of the results.

Threats to the Validity of the Primary Studies

Although experimental studies are the gold standard in separating causal from noncausal explanations, correlational studies can contribute to a causal explanation to the extent that they meet certain criteria (e.g., see Hill, 1965; Rothman & Greenland, 1998), the most important of which are temporality and control of third variables that could plausibly account for an association. Other threats include the use of retrospective reports or shared method variance in assessing parenting behavior and children's outcomes.

Establishing temporal sequences between risks and outcomes. A cross-sectional design cannot show that an association is a risk factor because it cannot establish temporal order. Fifty-eight percent of all the effect size estimates in Gershoff's (2002) meta-analyses came from cross-sectional analyses, and only 17% came from either experimental or prospective longitudinal studies. Thus, most of the primary studies did not provide adequate information for identifying a risk factor, much less a causal risk factor. A *causal risk factor* for an outcome is a risk factor believed to be a generative cause of that outcome—that is, to have produced that outcome. Experimental designs or systematic attempts to rule out plausible confounds in quasi-experimental and prospective longitudinal designs are necessary to establish a risk factor as causal.

Prospective longitudinal studies, unlike cross-sectional studies, can establish temporality and have more potential for controlling third variables adequately. They can provide evidence of causal direction from parents' use of CP to subsequent child behaviors by controlling for baseline child behaviors that provoke punishment and for the autocorrelation of behavior over time. We should note that Gershoff cited a paper by Cowan, Powell, and Cowan (1998) to the effect that correlational and follow-back studies are not sufficient to establish causality. Although they have some design advantages that we have highlighted in this section, it is still important to note that longitudinal models are necessary but not sufficient to establish the order of events. Parenting behavior assessed at Time 1 may predict child behavior at Time 2, but it is also possible that the parent may be reacting to child behavior that occurred before the study began.

Controlling for third variables. In their seminal discussion of the use of meta-analytic synthesis for theory development, Miller and Pollock (1994) stated "Assessing evidence that a relationship exists and that it is nonartifactual is a first step in assessing its validity" (p. 463). In correlational studies (as Straus, 1994, p. 217 conceded) this requires controlling for plausible third variables. For a relationship not to be spurious, it must persist in the form of a nonzero partial correlation between the supposed independent variable and the dependent variable with plausible third variables held constant (for example, baseline child misbehavior or parental rejection). Quasi-experimental studies are designed to exclude spurious variables from the group of causal influences affecting the results being observed. However, Gershoff (2002) used

DSTAT, whose manual explicitly prohibits basing effect sizes on coefficients from multiple regression (Johnson, 1989). If only the zero-order correlations can be used to calculate effect size, the methodological advantage of most quasi-experimental studies is lost, which makes the meta-analytic conclusions less relevant for determining whether CP is a causal risk factor for the detrimental child outcomes with which it is associated.

The use of retrospective recall. As Gershoff (2002) pointed out, reliance on retrospective recall in all but three of the studies linking CP in childhood to adult aggression, criminality, mental health, and physical abuse is a severe methodological inadequacy of those studies. Typically, retrospective recall is biased by the current status of the reporter. For example, adult reporters asked to report on their own depression and their parents' disciplinary practices may try to explain their current mental state by their parents' mistreatment.

Shared method variance. Problems associated with the fact that raters are not independent and thus that information about parenting and child outcomes is filtered through the same source constitute a critical limitation of half of the primary studies in the review. Shared method variance results in spuriously high correlations among constructs (Kenny & Kashy, 1992). The largest proportion of independent data sources occurred in compliance studies (four of five), whereas the smallest proportion of independent data sources were in adulthood studies of abuse (one of five) and aggression (one of four). The bias introduced by shared method variance is well illustrated in the analyses by Yarrow et al. (1968). They grouped their results in tables to demonstrate how a "contaminated design" artifactually increases an association. Mothers' reports on "severity of punishment for aggression to parents" correlated at .40 with their reports on the child's aggression toward adults in the home (which the authors refer to as the *contaminated design*), compared with a .13 correlation with teacher-reported aggression (Yarrow et al., 1968, pp. 72, 74). Gershoff (2002) chose to base her effect size on the correlation of .40 from the contaminated design rather than the .13 correlation in the uncontaminated design.

Threats to the Validity of the Synthesis

In addition to the problems caused by an overinclusive definition of CP, which could inflate effect sizes, Gershoff (2002) also based effect size estimates partly on cross-sectional correlations when longitudinal correlations were available (e.g., Sears, 1961; Straus, Sugarman, & Giles-Sims, 1997). An alternative strategy would have been to base effect sizes on the most valid associations available in a study or the ones most relevant to the debate about mild or moderate corporal punishment. Gershoff did not state the rules she used to include or exclude effect sizes when several options were available in a study. A rule that seems reasonable in selecting among relevant effect sizes within a study is to choose the one most likely to both exclude severe CP and to minimize shared method variance. To do otherwise would systematically increase the estimate of effect size with contaminated designs (Yarrow et al., 1968) and with parents who use CP too severely.

The Yarrow et al. study (1968) illustrates the somewhat arbitrary nature of the myriad of decisions the synthesist must make in selecting which statistic should represent a study. Two tables summarized correlations of four punishment measures with three

aggression measures. Gershoff based her effect size estimate on the largest of these 12 correlations. Furthermore, that correlation seems to be the poorest choice in terms of CP severity and validity threats. It was a correlation of .40 of "severity of punishment for aggression to parents" and aggression to adults at home, both based on the maternal interview. The most valid correlation was $-.19$ between maternal-reported physical punishment (overall) and teacher-reported aggression 2 months later. It was more valid because it was based on physical punishment overall, not severity of punishment in general, and because the CP and aggression measures were collected at different occasions from independent sources. Thus, the basis of an effect size between CP and child's aggression varies from $-.19$ in an uncontaminated design to .40 in a design contaminated by shared method variance, cross-sectional data, and overly severe punishment.

Furthermore, following the DSTAT manual's proscription against basing effect sizes on regression coefficients, even when controlling for third variables thought to be plausible confounds (Johnson, 1989), Gershoff's effect sizes never reflected the most causally relevant statistics in longitudinal designs. Other meta-analytic experts recommend controlling for plausible confounds whenever possible (Fleiss, 1994; Greenland, 1998), and some provide equations for calculating effect sizes from regression coefficients (Glass, McGaw, & Smith, 1981).

Reanalysis of the Data to Correct for Flaws in the Included Aggression Composite Studies

There are different views about whether poorly designed studies should be included in a meta-analysis. Mansfield and Busse (1977) recommended, and we concur, that all studies with severe methodological inadequacies should be eliminated from consideration and the remainder divided into two categories: those that have significant but not severe methodological limitations (e.g., retrospective longitudinal studies for which the likelihood of biased recall may be small) and those that are well designed in that they do not suffer from shared method variance, do use a prospective longitudinal or experimental design, and explicitly rule out severe corporal punishment. Unfortunately, as we show, that would leave too few studies for a meta-analytic review of the link between CP and most child behaviors (a situation in which one could conclude that a meta-analysis is premature).

Alternatively, methodological qualities of the studies can be coded and then used to see whether effect sizes from poorly designed studies are significantly different from those from the best-designed studies (Glass, 1976). Gershoff (2002) did attempt such analyses with her aggression composite variables, but these analyses were limited in two ways. First, almost all the studies had major methodological problems. More than half of the studies in the aggression composite used overly severe CP measures, more than half used cross-sectional or retrospective designs, and half had shared method variance. These problems were confounded with each other in unknown ways, making it difficult to sort out the differences those problems made by themselves or in combination. Second, Gershoff's analysis of study characteristics used a very conservative criterion for significance, especially considering the large number of categorical predictors (21) and small number of studies (22). This explains the unusual result of accounting for 86% of the variance in effect sizes (Gershoff, 2002, Table 6)

without any of the categorical predictors being significant according to her conservative Bonferroni correction.

In contrast, we implemented more realistic analyses of whether effect sizes varied by CP severity, research design, or independence of data sources. We analyzed Gershoff's effect sizes in her aggression composite using the chi-square analyses of weighted effect sizes recommended by Hedges (1994). Our analyses differ from Gershoff's analyses in the following ways: We treated each effect size as a separate case rather than combining effect sizes into one per study; we analyzed each methodological issue as a main effect rather than incorporating all predictors into one combined analysis; and finally, we used our codes for overly severe CP and corrected the design code for six cases.¹

As expected, effect sizes varied significantly by CP severity, by research design, and by independence of data sources. We found significantly higher effect sizes for overly severe than for nonsevere CP measures (mean weighted effect sizes, $d_+ = 0.46$ vs. 0.30 , respectively, $\chi^2[1, N = 12,244] = 74.5, p < .001$). Effect sizes were also higher for cross-sectional than for longitudinal studies ($d_+ = 0.46$ vs. 0.37 , $\chi^2[1, N = 7,807] = 16.1, p < .001$). Finally, effect sizes were higher for studies based partly or entirely on shared data sources than for independent sources of data ($d_+ = 0.35$ vs. 0.29 , $\chi^2[1, N = 13,591] = 8.44, p < .01$). Only 3 of 54 studies in the aggression composite were methodologically sound on all three study characteristics (Johannesson, 1974; McCord, 1988a, 1988b). Their mean weighted effect size was 0.12 , significantly less than the other 51 effect sizes in the aggression composite ($d_+ = 0.33$, $\chi^2[1, N = 13,591] = 5.99, p < .05$). This effect size is smaller than what Cohen (1988) labeled as small ($d = 0.20$).²

Gershoff's (2002) analyses were limited in investigating important substantive distinctions as well. For example, there have been some discussions recently about whether the associations of CP differ by ethnicity (e.g., Deater-Deckard & Dodge, 1997). Gershoff could not investigate ethnic differences because none of her primary studies investigated ethnic minorities alone. Nonetheless, two of her primary studies included separate statistics for African American subsamples (Deater-Deckard et al., 1996; Gunnoe & Mariner, 1997). In both studies, the mean effect sizes were negative, such that CP predicted less aggression subsequently at school in the African American subsample ($d = -0.18$ from Gunnoe & Mariner, 1997; $d = -0.17$ from Deater-Deckard et al., 1996). Furthermore, Gunnoe and Mariner (1997) went beyond a longitudinal correlation by controlling statistically for baseline aggression, which resulted in stronger causal evidence that CP reduces subsequent school aggression in African American children ($\beta = -.30$). At least two other studies have also found that the association of CP with antisocial behavior differs significantly for African American and European American children (McLeod, Kruttschnitt, & Dornfeld, 1994; Polaha, 1998).

Similarly, it is not clear that mild to moderate CP increases subsequent aggression and antisocial behavior at ages when parents are most likely to use CP (18 months to 6 years of age). Only one effect size in Gershoff's aggression construct was based on longitudinal and independent data on the use of nonsevere CP in this age range, and it was one of the smallest effect sizes ($d = 0.06$; Johannesson, 1974). Other primary studies had tiny effect sizes or ones for which CP predicted reductions in aggression for this age range when focusing on a young subsample (Gunnoe & Mariner,

1997) and using data uncontaminated by the above three methodological problems (Yarrow et al., 1968). Furthermore, Gunnoe and Mariner (1997) showed that CP predicted significantly less fighting in 4- to 7-year-olds after controlling for baseline antisocial behavior.

Comments Concerning Gershoff's Interpretation of the Meta-Analyses

Family Systemic Determinants of Children's Outcomes

Gershoff (2002) provided a detailed analysis of the possible mediators and moderators that could serve to amplify or protect against the purported risks of CP. She explored the potential role of child characteristics, parent-child interaction patterns, marital conflict, social support, and cultural forces in determining whether and how various parenting practices might have an adverse effect on children's development. However, she examined each of these variables one at a time. A view more consistent with a family-systems theory of parenting (e.g., Cowan et al., 1998) would be that each domain within and outside family life serves as a context that shapes the way in which any single practice affects the child's development.

CP and Physical Abuse

The high association between CP and physical abuse does not imply a causal link as Gershoff (2002) seemed to suggest when she said, "Child abuse in any form is a tragedy and deserves our best prevention efforts, and thus the potential for corporal punishment to escalate into physical abuse must be seriously considered at the levels of scientific research and public policy" (p. 550). The relatively large effect size of the association between CP and physical abuse is not evidence that mild or moderate CP increases the risk of abuse. Similar unwarranted causal inferences have been made that marijuana experimentation necessarily increases the risk of heroin abuse (Baumrind, 1983b). In both cases the phenomena are correlated but quite distinct from each other.

If both CP and physical abuse are measured with overlapping levels of severity, the conclusion that they are closely linked is tautologous. By including severe and excessive levels of hitting in her operational definition of CP, Gershoff lost the conceptual refinement that could reveal a qualitative distinction or threshold phenomenon that would transform a quantitative difference into a qualitative one, much as increases in speed result in qualitatively different gaits in a horse as it moves from a walk (slow, four-beat gait with feet striking the ground in a specific order) to a trot (legs move in diagonal pairs) to a canter or gallop (three-beat gait); the qualitatively different gaits are not reducible to the single parameter of speed.

¹ Two studies were recoded as cross-sectional because Gershoff's (2002) effect sizes were based on cross-sectional data from longitudinal (Stattin, Janson, Klackenber-Larsson, & Magnusson, 1995) or retrospective (Flynn, 1999) studies. Four effect sizes were recoded as longitudinal because they were apparently miscoded as cross-sectional (Simons et al., 1994, 2 effect sizes; Simons, Lin, & Gordon, 1998; Straus et al., 1997).

² Equivalent correlations are about half of the mean d values in this paragraph, thus ranging from .06 to .23.

Parents who escalate to abuse from disciplinary use of CP are likely to share a distinctive set of attributes (Baumrind, 1995; Vasta, 1982). Abusive parents are more likely to be hyperreactive to negative stimuli and to have an extreme need to control their children. Their punishment is less contingent on the child's behavior than on their own inner state. Rather than having flexible recourse to a wide range of disciplinary tactics, such as time-out, induction, persuasion, and denial of privileges, abusive parents rely monolithically on their greater physical power to intimidate their child into compliance. Their anger is explosive, and they hit impulsively in response to their own frustration rather than to correct the child. As Gershoff (2002) stated, it is important for future research to determine for whom, in what contexts, and when purposeful "corporal punishment is transformed into abuse" (p. 553). As her term implies, a qualitative change must occur for ordinary physical punishment to be "transformed" into physical abuse. Thus, in a study of affluent, well-educated families, those parents whose recourse to physical punishment was excessively severe and frequent also engaged in significantly more negative interactions of other kinds including verbal abuse, being significantly less warm, supportive and consistent, and themselves exhibiting more internalizing and externalizing problem behavior (Baumrind, 2001).

Just as addictive personalities should not drink alcohol or use drugs, some parents—those with a low tolerance for frustration, a history of violence, an inordinate need to control others, and those who are impulsive, narcissistic, and immature—should not spank. The fact that some parents punish excessively and unwisely is not an argument, however, for counseling all parents to not punish at all, anymore than the fact that some people drink excessively is a reason to counsel against all alcohol consumption.

Are the Links Between Adverse Child Outcomes and CP Unique?

Gershoff (2002) stated, "These meta-analyses focused on corporal punishment, and their findings should not be extended unequivocally to other forms of punishment, such as time-out or withdrawal of privileges" (p. 551). We examined the effect sizes of the associations between other forms of punishment and aggression composite scores in the small subset of primary studies containing that information and found that the effect sizes suggested more detrimental links for alternative forms of punishment in younger children.

Seven studies in Gershoff's aggression composite used the same methods to investigate alternative punishment tactics as they did for CP. Four were studies of teenagers, which generally found less detrimental effect sizes associated with alternative punishment tactics than with CP (Bandura & Walters, 1959; Glueck & Glueck, 1950), especially in the two studies that forced respondents to choose one major disciplinary method (Caesar, 1988; Kahn & Fua, 1995). This confirms the generally accepted view that physical correction is inappropriate with teenagers. In contrast, the three studies of younger children all found more detrimental associations (effect sizes) for alternative punishment tactics than for physical punishment (Sears, 1961; Straus & Mouradian, 1998; Yarrow et al., 1968).

Because the effect sizes linking CP to detrimental outcomes in young children are often smaller than those linking other forms of

punishment to such outcomes, we might conclude that parents should refrain from all forms of punishment because all punishment harms young children. Alternatively, the links between CP (or other forms of punishment) and detrimental outcomes in young children may be *artifactual* (i.e., due in part or wholly to third variables that were not controlled for, such as baseline child misbehavior and family context).

Behavioral Compliance as a Beneficial Outcome

The reason nonabusive parents use punishment, including CP, is primarily to achieve short-term *behavioral compliance* (i.e., adapting one's actions to conform with the direction of others), which should be distinguished from *dispositional compliance* (Baumrind, 1983a). A person who internalizes the general norm of compliance to authority may be said to be dispositionally compliant. High dispositional compliance is viewed ambivalently in Western cultures, which place an important value on autonomy and individuality. A child who is not dispositionally compliant, however, is likely also to be less malleable and therefore likely to require more forceful parental intervention to secure behavioral compliance.

The distinction between behavioral compliance and dispositional compliance has much in common with Kochanska, Coy, and Murray's (2001) distinction between *committed compliance* (when children eagerly embrace the caregiver's agenda) and *situational compliance* (when they cooperate for instrumental reasons, but without a commitment). Committed compliance in the "don't" condition, in which the child was required to suppress prohibited behavior, was associated positively with children's fearfulness and shyness (Kochanska et al., 2001). Fearful children are more easily conditioned to inhibit transgression than bold children (Dienstbier, 1984), who are more likely to defy parental authority. Punishment, in particular physical punishment, is not only functionally superfluous for shy, fearful children (Lepper, 1981) but may be traumatic.

Debates about the importance of short-term compliance may depend on the type of noncompliance. Clinical psychologists focus on the defiant end of the continuum, whereas other psychologists focus on more functional types and levels of noncompliance (Kuczynski & Hildebrandt, 1997; Kuczynski & Kochanska, 1990). From both perspectives, a moderate level of compliance is optimal, and some types of noncompliance are more functional developmentally than others. Defiant noncompliance in young children predicts poorer moral internalization as well as a greater risk of antisocial behavior (Kochanska & Aksan, 1995; Loeber & Schmalting, 1985; Lytton, Watts, & Dunn, 1986; Patterson, Reid, & Dishion, 1992). Mothers are also more likely to respond to defiant noncompliance with harsh forms of punishment (Ritchie, 1999).

Interventions as a Way of Understanding the Linkage Between CP and Children's Behavior

Participants in the debate on CP often point out that it would be unethical to test causal hypotheses by randomly assigning children to experimental and control groups. However, it is also the case that intervention designs can shed light on the issue of how parents' discipline styles affect their children. Behavioral parent training is a clinical treatment that has decreased a wide range of problem behaviors in referred children and increased their appro-

ropriate behaviors—improvements that have been maintained as long as 14 years. Indeed, a recent meta-analysis of 26 comparison group studies (18 randomized) found a mean effect size (d) for behavioral parent training of -0.85 for observational reports of child antisocial behavior, -0.73 for teacher reports of antisocial behavior, and 0.44 for improved parental adjustment (e.g., marital satisfaction and lower depression or stress; Serketich & Dumas, 1996). Not all 26 studies used a spank back-up for time-out in their version of behavioral parent training. However, the 5 studies in that meta-analysis that we could locate that did use a spank back-up for time-out all reported significantly greater improvement on either aggression or behavior problems relative to a comparison group (Hamilton & MacQuiddy, 1984; McNeil, Eyberg, Eisenstadt, Newcomb, & Funderburk, 1991; Olson & Roberts, 1987; Peed, Roberts, & Forehand, 1977; Zangwill, 1983).

Other studies have documented longer term maintenance of improvements from behavioral parent training, including versions with a spank back-up. Six of the 26 studies reviewed in the Serketich and Dumas (1996) meta-analysis documented that improvements were maintained at follow-up, which varied from 2 to 12 months after treatment. Long, Forehand, Wierson, and Morgan (1994) found that clinical referrals treated with behavioral parent training with a spank back-up at ages 2 to 7 were indistinguishable from matched community controls 14 years later on relationship quality with parents, delinquency, substance abuse, and emotional adjustment. Baum and Forehand (1981) showed that posttreatment gains were maintained from 1 to 4.5 years after the same type of behavioral parent training.

Behavioral parent trainers see the task of improving compliance to normal levels as a crucial initial step for decreasing other forms of antisocial behavior (Barkley, 1981; Loeber & Schmalting, 1985; Lorber & Patterson, 1981). From this perspective, children must improve their compliance to parents to normal levels before parents can begin to have a positive influence on increasing their prosocial behavior and decreasing their referral problems (e.g., aggression, noncompliance with medical regimens). Compliance must be improved first in the clinic before it can be improved at home. There is no evidence that spanking is either a more or a less effective deterrent than other punishment methods with young children, although time-out with a barrier or restraint back-up is the alternative currently favored by most behavioral parent trainers for defiant children. However, mothers of defiant 2- to 6-year-old children in a study by Roberts and Powers (1990) preferred a “two-swat” back-up rather than a barrier or restraint back-up to enforce time-out. It remains to be studied whether parents can and will use an alternative back-up such as a barrier with a defiant child, especially in homes where space and time are limited.

Conclusion

Gershoff (2002) acknowledged the necessity to distinguish between the harsh, punitive parental discipline that all experts would regard as both harmful and unethical (including such advocates of spanking as Dobson, 1992, and Rosemond, 1994) and the more normative parental actions that involve the infliction of mild physical pain but not injury. In this response to Gershoff’s article, we reanalyzed portions of Gershoff’s meta-analyses to show that this separation was not in fact made in many primary studies.

Because her measure included many instances of extreme and excessive physical punishment, her analyses are not relevant to the current political debate about whether normative spanking (a specific form of CP) is harmful for children.

Punishment by definition is aversive and therefore must be used sparingly and efficaciously to be acceptable. As Larzelere (2000) stated, “There may be more potential in identifying more vs. less effective ways to use each disciplinary tactic. The effect of any tactic may depend on the overall disciplinary style and the parent-child, family, and cultural contexts” (p. 217). Larzelere (2001) suggested a conditional sequence model as optimal in which parents begin with the mildest disciplinary tactic they think will be effective, such as reasoning, and resort to firmer tactics only if the child is defiant about a nonnegotiable request. If used efficaciously as a back-up at ages 2 to 6, both physical and nonphysical punishment are less likely to be needed as children learn to attain a reasonable level of behavioral compliance to mild disciplinary tactics such as reasoning.

Larzelere’s (2001) conditional sequence model and Gershoff’s (2002) process-context model should be tested with well-designed studies from culturally diverse populations. Even though randomized clinical trials are probably not practical, intervention studies are possible that would add to the knowledge base. To determine if normative physical punishment has any unique effect on 2- to 6-year-olds over and above that of otherwise optimal parenting, parents could be trained in techniques of authoritative discipline, which would include efficacious use of punishments of their choice. Some parents would choose to include spanking in their disciplinary repertoire, providing the opportunity for researchers to assess the child outcomes associated with disciplinary spanking when used efficaciously by comparing them with child outcomes associated with the alternative disciplinary tactics used by parents who choose to never spank. With a large and diverse enough sample, it should be possible to evaluate the moderating effects of third variables that would not be controlled, such as child and parent temperament and social stressors, in such quasi-experimental intervention studies.

Gershoff’s (2002) own conceptual analysis of potential mediators and moderators affecting the linkage between CP and children’s outcomes underscored how simplistic most CP research has been to date. What is needed in research on parenting are the kinds of studies implied by Gershoff’s analysis in the second section of her article that ask: In what contexts is what kind of parental discipline likely to (a) increase the probability of positive outcomes and decrease the probability of negative ones and (b) increase the probability of negative outcomes and decrease the probability of positive ones? Any further studies of the links between corporal punishment and child outcomes should be limited to a range of frequency and severity that some spokespersons regard as beneficial relative to its costs—namely, mild to moderate well-regulated spanking on the buttocks or extremities in children over 18 months and prior to adolescence. To date, the long-term outcomes associated with a voluntary decision by parents to never use CP, or indeed any form of punishment (McCord, 1991), has not been studied systematically. At present we conclude that the evidence presented in Gershoff’s meta-analyses does not justify a blanket injunction against mild to moderate disciplinary spanking.

References

- American Academy of Pediatrics. (1998). Guidance for effective discipline. *Pediatrics*, *101*, 723–728.
- Bandura, A., & Walters, R. H. (1959). *Adolescent aggression*. New York: Ronald Press.
- Barkley, R. A. (1981). *Hyperactive children: A handbook for diagnosis and treatment*. New York: Guilford Press.
- Baum, C. G., & Forehand, R. (1981). Long term follow-up assessment of parent training by use of multiple outcome measures. *Behavior Therapy*, *12*, 643–652.
- Baumrind, D. (1983a). Rejoinder to Lewis's reinterpretation of parental firm control effects: Are authoritative parents really harmonious? *Psychological Bulletin*, *94*, 132–142.
- Baumrind, D. (1983b). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. *Journal of Personality and Social Psychology*, *45*, 1289–1298.
- Baumrind, D. (1995). *Child maltreatment and optimal caregiving in social contexts*. New York: Garland.
- Baumrind, D. (1996a). A blanket injunction against disciplinary use of spanking is not warranted by the data. *Pediatrics*, *98*, 828–831.
- Baumrind, D. (1996b). The discipline controversy revisited. *Family Relations*, *45*, 405–414.
- Baumrind, D. (2001, August). *Does causally relevant research support a blanket injunction against disciplinary spanking by parents?* Invited address at the 109th Annual Convention of the American Psychological Association, San Francisco, CA.
- Bitensky, S. H. (1997). Spare the rod, embrace our humanity: Toward a new legal regime prohibiting corporal punishment of children. *University of Michigan Journal of Law Reform*, *31*(1), 353–474.
- Caesar, P. L. (1988). Exposure to violence in the families-of-origin among wife-abusers and maritally nonviolent men. *Violence and Victims*, *3*, 49–63.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally.
- Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., et al. (1992). *Meta-analysis for explanation: A casebook*. New York: Sage.
- Cowan, P. A., Powell, D., & Cowan, C. P. (1998). Parenting interventions: A family systems perspective. In W. Damon (Series Ed.) & I. E. Sigel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed., pp. 3–72). New York: Wiley.
- Deater-Deckard, K., & Dodge, K. A. (1997). Externalizing behavior problems and discipline revisited: Nonlinear effects and variation by culture, context, and gender. *Psychological Inquiry*, *8*, 161–175.
- Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1996). Physical discipline among African American and European American mothers: Links to children's externalizing behaviors. *Developmental Psychology*, *32*, 1065–1072.
- Dienstbier, R. A. (1984). The role of emotion in moral socialization. In C. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognitions, and behaviors* (pp. 484–513). New York: Cambridge University Press.
- Dobson, J. (1992). *The new dare to discipline*. Wheaton, IL: Tyndale House.
- Engfer, A., & Schneewind, K. A. (1982). Causes and consequences of harsh parental punishment. *Child Abuse & Neglect*, *6*, 129–139.
- EPOCH-USA. (2000). *Mission statement*. Retrieved April 16, 2002, from <http://www.stophitting.com/disathome/>
- Eron, L. D. (1982). Parent-child interaction, television violence, and aggression of children. *American Psychologist*, *37*, 197–211.
- Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 245–260). New York: Sage.
- Flynn, C. P. (1999). Exploring the link between corporal punishment and children's cruelty to animals. *Journal of Marriage and the Family*, *61*, 971–981.
- Friedman, S., & Schonberg, S. K. (1996). Consensus statements. *Pediatrics*, *98*, 853.
- Garbarino, J. (1996). Invited commentary: CAN reflections on 20 years of searching. *Child Abuse & Neglect*, *20*, 157–160.
- Gershoff, E. T. (2002). Parental corporal punishment and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, *128*, 539–579.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Glueck, S., & Glueck, E. (1950). *Unraveling juvenile delinquency*. Cambridge, MA: Harvard University Press.
- Gordon, J. E., & Smith, E. (1965). Children's aggression, parental attitudes, and the effects of an affiliation-arousing story. *Journal of Personality and Social Psychology*, *1*, 654–659.
- Greenland, S. (1998). Meta-analysis. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 643–673). Philadelphia: Lippincott-Raven.
- Gunnoe, M. L., & Mariner, C. L. (1997). Toward a developmental-contextual model of the effects of parental spanking on children's aggression. *Archives of Pediatrics and Adolescent Medicine*, *151*, 768–775.
- Hamilton, S. B., & MacQuiddy, S. L. (1984). Self-administered behavioral parent training: Enhancement of treatment efficacy using a time-out signal seat. *Journal of Clinical Child Psychology*, *13*, 61–69.
- Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 285–300). New York: Sage.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, *58*, 295–300.
- Holden, G. W., Thompson, E. E., Zambardino, R. J., & Marshall, L. A. (1997). Child effects as a source of change in maternal attitudes toward corporal punishment. *Journal of Social and Personal Relationships*, *14*, 481–490.
- Holmes, S. J., & Robins, L. N. (1988). The role of parental disciplinary practices in the development of depression and alcoholism. *Psychiatry*, *51*, 24–35.
- Hyman, I. A. (1996). Using research to change policy: Reflections on 20 years of effort to eliminate corporal punishment in the schools. *Pediatrics*, *98*, 818–821.
- Johannesson, I. (1974). Aggressive behavior among school children related to maternal practices in early childhood. In J. De Wit & W. W. Hartup (Eds.), *Determinants and origins of aggressive behavior* (pp. 413–426). The Hague, the Netherlands: Mouton.
- Johnson, B. T. (1989). DSTAT: Software for the meta-analytic review of research literatures. Hillsdale, NJ: Erlbaum.
- Kahn, M. W., & Fua, C. (1995). Children of South Sea Island immigrants to Australia: Factors associated with adjustment problems. *International Journal of Social Psychiatry*, *41*, 55–73.
- Kandel, D. B. (1990). Parenting styles, drug use, and children's adjustment in families of young adults. *Journal of Marriage and the Family*, *52*, 183–196.
- Kazdin, A. E., Kraemer, H. C., Kessler, R. C., Kupfer, D. J., & Offord, D. R. (1997). Contributions of risk-factor research to developmental psychopathology. *Clinical Psychology Review*, *17*, 375–406.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.
- Kochanska, G., & Aksan, N. (1995). Mother-child mutually positive

- affect, the quality of child compliance to requests and prohibitions, and maternal control as correlates of early internalization. *Child Development*, 66, 236–254.
- Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development*, 72, 1091–1111.
- Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, 54, 337–343.
- Kuczynski, L., & Hildebrandt, N. (1997). Models of conformity and resistance in socialization theory. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and the internalization of values: A handbook of contemporary theory* (pp. 227–256). New York: Wiley.
- Kuczynski, L., & Kochanska, G. (1990). Development of children's non-compliance strategies from toddlerhood to age 5. *Developmental Psychology*, 26, 398–408.
- Larzelere, R. E. (2000). Child outcomes of non-abusive and customary physical punishment by parents: An updated literature review. *Clinical Child and Family Psychology Review*, 3, 199–221.
- Larzelere, R. E. (2001). Combining love and limits in authoritative parenting. In J. C. Westman (Ed.), *Parenthood in America* (pp. 81–89). Madison: University of Wisconsin Press.
- Lefkowitz, M. M., Walder, L. O., & Eron, L. D. (1963). Punishment, identification and aggression. *Merrill-Palmer Quarterly*, 9, 159–174.
- Lepper, M. (1981). Intrinsic and extrinsic motivation in children: Detrimental effects of superfluous social controls. In W. A. Collins (Ed.), *Aspects of the development of competence: Vol. 14. The Minnesota Symposium on Child Psychology* (pp. 155–214). Hillsdale, NJ: Erlbaum.
- Lester, D. (1991). Physical abuse and physical punishment as precursors of suicidal behavior. *Stress Medicine*, 7, 255–256.
- Levin, H., & Sears, R. R. (1956). Identification with parents as a determinant of doll play aggression. *Child Development*, 27, 135–153.
- Loeber, R., & Schmalting, K. B. (1985). Empirical evidence for overt and covert patterns of antisocial conduct problems: A meta-analysis. *Journal of Abnormal Child Psychology*, 13, 337–352.
- Long, P., Forehand, R., Wierson, M., & Morgan, A. (1994). Does parent training with young noncompliant children have long-term effects? *Behaviour Research and Therapy*, 32, 101–107.
- Lorber, R., & Patterson, G. R. (1981). The aggressive child: A concomitant of a coercive system. In J. P. Vincent (Ed.), *Advances in family intervention, assessment and theory: An annual compilation of research* (pp. 47–87). Greenwich, CT: JAI Press.
- Lytton, H., Watts, D., & Dunn, B. E. (1986). Stability and predictability of cognitive and social characteristics from age 2 to age 9. *Genetic Social and General Psychology Monographs*, 112, 363–398.
- Mahoney, A., Donnelly, W. O., Lewis, T., & Maynard, C. (2000). Mother and father self-reports of corporal punishment and severe physical aggression toward clinic-referred youth. *Journal of Clinical Child Psychology*, 29, 266–281.
- Mann, C. C. (1994, November 11). Can meta-analysis make policy? *Science*, 266, 960–992.
- Mansfield, R. S., & Busse, T. V. (1977). Meta-analysis of research: A rejoinder to Glass [Letter to the editor]. *Educational Researcher*, 6(9), 3.
- McCord, J. (1988a). Parental aggressiveness and physical punishment in long-term perspective. In G. T. Hotaling, D. Finkelhor, J. Kilpatrick, & M. A. Straus (Eds.), *Family abuse and its consequences* (pp. 91–98). Newbury Park, CA: Sage.
- McCord, J. (1988b). Parental behavior in the cycle of aggression. *Psychiatry*, 51, 14–23.
- McCord, J. (1991). Questioning the value of punishment. *Social Problems*, 38, 167–179.
- McLeod, J. D., Kruttschnitt, C., & Dornfeld, M. (1994). Does parenting explain the effects of structural conditions on children's antisocial behavior? A comparison of Blacks and Whites. *Social Forces*, 73, 575–604.
- McNeil, C. B., Eyberg, S., Eisenstadt, T. H., Newcomb, K., & Funderburk, B. (1991). Parent-child interaction therapy with behavior problem children: Generalization of treatment effects to the school setting. *Journal of Clinical Child Psychology*, 20, 140–151.
- Miller, N., & Pollock, V. E. (1994). Meta-analytic synthesis for theory development. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 457–484). New York: Russell Sage Foundation.
- Mrazek, P. J., & Haggerty, R. J. (1994). *Reducing risks for mental disorders: Frontiers for prevention intervention research*. Washington, DC: National Academy Press.
- Muller, R. T. (1996). Family aggressiveness factors in the prediction of corporal punishment: Reciprocal effects and the impact of observer perspective. *Journal of Family Psychology*, 10, 474–489.
- Olson, R. L., & Roberts, M. W. (1987). Alternative treatments for sibling aggression. *Behavior Therapy*, 18, 243–250.
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia.
- Peed, S., Roberts, M., & Forehand, R. (1977). Evaluation of the effectiveness of a standardized parent training program in altering the interaction of mothers and their noncompliant children. *Behavior Modification*, 1, 332–350.
- Polaha, J. A. (1998). *The relationship between physical discipline and child behavior problems: A study of group differences*. Unpublished doctoral dissertation, Auburn University, Auburn, AL.
- Riggs, D. S., & O'Leary, K. D. (1996). Aggression between heterosexual dating partners: An examination of a causal model of courtship aggression. *Journal of Interpersonal Violence*, 11, 519–540.
- Ritchie, K. L. (1999). Maternal behaviors and cognitions during discipline episodes: A comparison of power bouts and single acts of noncompliance. *Developmental Psychology*, 35, 580–589.
- Roberts, M. W., & Powers, S. W. (1990). Adjusting chair timeout enforcement procedures for oppositional children. *Behavior Therapy*, 21, 257–271.
- Rosemond, J. K. (1994). *To spank or not to spank*. Kansas City, MO: Andrews & McMeel.
- Rothman, K. J., & Greenland, S. (1998). Causation and causal inference. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 7–28). Philadelphia: Lippincott-Raven.
- Sears, R. R. (1961). Relation of early socialization experiences to aggression in middle childhood. *Journal of Abnormal and Social Psychology*, 63, 466–492.
- Sears, R. R., Maccoby, E. E., & Levin, H. (1957). *Patterns of child-rearing*. New York: Harper & Row.
- Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta-analysis. *Behavior Therapy*, 27, 171–186.
- Simons, R. L., Johnson, C., Beaman, J., & Conger, R. D. (1993). Explaining women's double jeopardy: Factors that mediate the association between harsh treatment as a child and violence by a husband. *Journal of Marriage and the Family*, 55, 713–723.
- Simons, R. L., Johnson, C., & Conger, R. D. (1994). Harsh corporal punishment versus quality of parental involvement as an explanation of adolescent maladjustment. *Journal of Marriage and the Family*, 56, 591–607.
- Simons, R. L., Lin, K.-H., & Gordon, L. C. (1998). Socialization in the family of origin and male dating violence: A prospective study. *Journal of Marriage and the Family*, 60, 467–478.
- Stattin, H., Janson, H., Klackenber-Larsson, I., & Magnusson, D. (1995). Corporal punishment in everyday life: An intergenerational perspective. In J. McCord (Ed.), *Coercion and punishment in long-term perspectives* (pp. 315–347). Cambridge, England: Cambridge University Press.
- Straus, M. A. (1990). Ordinary violence, child abuse, and wife beating:

What do they have in common? In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families* (pp. 403–424). New Brunswick, NJ: Transaction.

Straus, M. A. (1994). *Beating the devil out of them: Corporal punishment in American families*. New York: Lexington Books.

Straus, M. A. (1999). Is it time to ban corporal punishment of children? *Canadian Medical Association Journal*, 161, 821–822.

Straus, M. A., & Mouradian, V. E. (1998). Impulsive corporal punishment by mothers and antisocial behavior and impulsiveness of children. *Behavioral Sciences and the Law*, 16, 353–374.

Straus, M. A., Sugarman, D. B., & Giles-Sims, J. (1997). Spanking by

parents and subsequent antisocial behavior of children. *Archives of Pediatrics and Adolescent Medicine*, 151, 761–767.

Vasta, R. (1982). Physical child abuse: A dual-component analysis. *Developmental Review*, 2, 125–149.

Yarrow, M. R., Campbell, J. D., & Burton, R. V. (1968). *Child rearing: An inquiry into research and methods*. San Francisco: Jossey-Bass.

Zangwill, W. M. (1983). An evaluation of a parent training program. *Child & Family Behavior Therapy*, 5(4), 1–16.

Received November 16, 2001
 Accepted November 16, 2001 ■



**AMERICAN PSYCHOLOGICAL ASSOCIATION
 SUBSCRIPTION CLAIMS INFORMATION**

Today's Date: _____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do NOT duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION _____		MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL) _____
ADDRESS _____		DATE YOUR ORDER WAS MAILED (OR PHONED) _____
CITY _____ STATE/COUNTRY _____ ZIP _____		<input type="checkbox"/> PREPAID <input type="checkbox"/> CHECK <input type="checkbox"/> CHARGE CHECK/CARD CLEARED DATE: _____
YOUR NAME AND PHONE NUMBER _____		(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.) ISSUES: <input type="checkbox"/> MISSING <input type="checkbox"/> DAMAGED
TITLE	VOLUME OR YEAR	NUMBER OR MONTH
_____	_____	_____
_____	_____	_____
_____	_____	_____

Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4–6 weeks.

(TO BE FILLED OUT BY APA STAFF)

DATE RECEIVED: _____	DATE OF ACTION: _____
ACTION TAKEN: _____	INV. NO. & DATE: _____
STAFF NAME: _____	LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.