Pergamon

Archives of Clinical Neuropsychology 16 (2001) 303–341

Archives of CLINICAL NEUROPSYCHOLOGY

Do low levels of lead produce IQ loss in children? A careful examination of the literature☆

Alan S. Kaufman

Child Study Center, School of Medicine, Yale University, 230 South Frontage Road, PO Box 207900, New Haven, CT 06520-7900, USA

Accepted 28 August 2000

Abstract

Three meta-analyses on the relationship of low levels of lead to loss of IQ points in children, which included a total of 26 well-controlled studies, provided the raw materials for the analysis presented here. Despite some key limitations, results of lead–IQ studies have been instrumental in setting public policy. In this paper, five shortcomings in these studies are addressed, which, when taken together, suggest greater caution in the interpretation of the lead–IQ data. In addition, some other issues are addressed concerning the IQ loss attributed to low levels of lead. © 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

Keywords: Blood lead level; IQ; Neurotoxins; Neuropsychological functioning; Lead research; Intelligence

Research on the relationship between the presence of low levels of blood lead (or bone lead or tooth lead) in children, and a decrease in their cognitive function, has been a source of controversy within the research literature; some researchers have argued that even very low levels of lead can have severe consequences on children's intellectual and academic functioning (Bellinger, Stiles, & Needleman, 1992) and other researchers have maintained that the effect size of lead on children's neuropsychological dysfunction is small or nonexistent (Ernhart, 1992, 1995; Ernhart, Morrow-Tlucak, Wolf, Super, & Drotar, 1989). Low blood lead levels (BLLs) in children are regarded as falling within the general boundaries of 10 to 20 μg/dl. According to the Centers for Disease Control's guidelines,

☆ A preliminary draft of portions of this manuscript was partially funded by Ethyl, Richmond, VA. The author gratefully acknowledges Ethyl for access to their library of published scientific articles, and recognizes Jerry M. Roper, Ph. D., of the same firm, for his assistance with the scientific literature and for helpful support and input on this project.

PII: S0887-6177(00)00084-6

BLLs below 10 μg/dl are interpreted as "safe," while medical evaluation and, in some cases, treatment are recommended for blood leads above 20 μg/dl (Centers for Disease Control, 1991). A BLL at or above 70 μg/dl is considered a medical emergency, and there is evidence that acute lead poisoning (> 80 μg/dl) can produce damage to brain tissue (Goldstein, 1984). The current paper does not deal with these higher BLLs, but rather addresses the question of what impact — if any — do low BLLs (10–20 μg/dl) have on children's performance on conventional measures of intellectual functioning. Not addressed here is the more general topic of how BLL affects general neuropsychological functioning. Consistent with the main thrust of the best-designed and best-controlled studies of the relationship of BLL to children's neuropsychological integrity (e.g., Schwartz, 1994), this paper focuses on conventional IQ tests and asks whether the literature supports a significant relationship between low BLLs and IQ loss.

Science, especially where results and conclusions are vigorously debated, does not easily lend itself to the proposition of public policy. Concerning the debate over the effect of low BLLs and IQ, those who set public policy favor the argument claiming a noteworthy effect size for low BLLs. For example, the concept that low BLLs produce measurable IQ loss has formed the basis for proposed comprehensive and expensive public policies, such as those associated with the implementation of various legislation (e.g., Residential Lead-Based Paint Hazard Reduction Act of 1992, 1992, Title X). Acceptance of the conclusions about the effects of low BLLs has led to agencies like the Department of Housing and Urban Development (HUD) to make statements such as: "more recent meta-analysis estimated .257 IQ points lost per one μg/dl increase in blood lead levels" (Federal Register, 1996, p. 29201). The latter statement infers a linear relationship between BLL and IQ loss and infers that a single IQ point can be meaningfully fragmented into tenths of a point. Both inferences need to be evaluated scientifically. In addition, some researchers (e.g., Needleman, 1989) have emphasized the potential societal impact of IQ loss due to exposure to low BLLs in the environment, such that an IQ loss hypothesized to average about 6 points for groups "results in a fourfold increase in the rate of severe deficit (IQ < 80) ... [and] that 5 per cent of lead-exposed children are prevented from achieving truly superior function (IQ>125)" (p. 643).

Additionally, elevated BLL has been implicated as a causative factor of attentional disorders and behavioral problems (Needleman et al., 1979; Yule, Urbanowicz, Lansdown, & Millar, 1984), a cause–effect proposition that has received much attention, especially regarding juvenile delinquency (Needleman, Riess, Tobin, Biesecker, & Greenhouse, 1996). Nonetheless, in this paper, I focus primarily on the relationship of low BLLs to alleged loss of intellectual function in children. Though attempts to relate lead to anti-social behavior and disorders such as Attention-Deficit Hyperactivity Disorder (ADHD) are provocative, the research basis for these contentions is as yet meager relative to the extensive epidemiological and experimental literature on lead and IQ (see, for example, Sachs, 1996; Sayre, 1996; Wasserman, Staghezza-Jaramillo, Shrout, Popovac, & Graziano, 1998).

The relationship of low BLLs in children to IQ is explored in two main sections. First, I will delineate five general shortcomings in the methodology associated with some of the best-designed studies that have attempted to relate low BLLs to IQ loss. Then, in a second section, I will deal with some questionable applications of the IQ loss attributed to children's lead level. Specifically, I will examine the inference of a steady loss of IQ with each small incremental increase in BLL, the purported societal impact of the loss of a few IQ points, and the practice of assigning significance to a single IQ point that has been fragmented into tenths or hundredths of a point.

1. Critique of the best-designed studies relating low BLLS to IQ loss

The initial research studies relating BLL to IQ failed to control for covariates like socioeconomic and familial factors, making it difficult to interpret the accuracy of the findings (e.g., Kotok, 1972; Landrigan, Balow, Whitworth, Staeling, & Rosenbloom, 1975). Subsequent studies employed improved methodology, and these "better" studies are the subject of this paper. I will define as the "better" studies the 26 investigations that were included in one or more of three meta-analyses: Needleman and Gatsonis (1990); Pocock, Smith, and Baghurst (1994); and Schwartz (1994). Each meta-analysis limited the investigations to the ones that met specified criteria of quality, primarily in terms of controlling for pertinent confounding covariates. Age ranges differed to some extent in the three meta-analyses: (a) Needleman and Gatsonis (1990) included studies of children of all ages, the youngest being 1 year and the oldest 12 years; (b) Schwartz (1994) limited his analysis to studies of "school-age children"; and (c) Pocock et al. (1994) included studies of children ages 5 years and older. Children as old as 14 are included in a few studies, but adolescents are not included in most samples. All of the meta-analyses concluded that when the data from the best studies are merged, low BLLs are associated with IQ loss. Needleman and Gatsonis (1990) concluded from their meta-analysis of seven blood-lead and five tooth-lead studies, all of which used multiple regression analysis: "The hypothesis that lead impairs children's IQ at low dose is strongly supported by this quantitative review" (p. 673), although they do not quantify the IQ loss. In other writings, Needleman and his colleagues (e.g., Needleman, 1989) often attribute to lead an IQ loss of about 4 to 7 IQ points. Schwartz (1994) concluded from his meta-analysis of five cross-sectional and three longitudinal investigations that, "An increase in blood lead from 10 to 20 μg/dl was associated with a decrease of 2.6 IQ points in the meta-analysis" (p. 42). Finally, Pocock et al. concluded from 14 cross-sectional studies of blood lead and seven cross-sectional studies of tooth lead: "Overall synthesis of this evidence, including a meta-analysis, indicates that a typical doubling of body lead burden (from 10 to 20 μg/dl ...) is associated with a mean deficit in full-scale IQ of around 1–2 IQ points" (p. 1189). Perhaps the best overview of these meta-analyses is that low BLLs are associated with an IQ loss of an average of about 3 points, the same approximate magnitude of loss observed in three well-designed studies published subsequent to the meta-analyses (Tong, Baghurst, McMichael, Sawyer, & Mudge, 1996; Wasserman et al., 1994, 1997). Three points (0.2 of a standard deviation), however, is a small effect size. McLean (1995) states that when effect sizes are based on standard deviation (S.D.) units, "An effect size of less than 0.50 is small and suggests that the difference is probably not meaningful" (p. 39). Nonetheless, Rosenthal's (1990, 1991) thoughtful discussion of the potential value of small effect sizes in a practical sense should not be overlooked. Using the example of the effectiveness of aspirin in preventing heart attacks, Rosenthal notes that even a near-zero correlation of .04 "does not appear to be quite so small, especially if we can count ourselves among the four per 100 who manage to survive" (p. 775). Similarly, the small effect size of lead–IQ studies does not minimize the potential practical value of the results, so long as scientific evidence supports the validity of the approximately 3 points of IQ loss attributed to low BLLs. It is the validity of that body of scientific evidence that I am evaluating in this paper.

In all, a total of 26 different studies were included in the three meta-analyses. Each of these studies, whether they implicate low BLL as an IQ-depressing variable or whether they indicate that low BLL is non-significant as a contributor to intellectual functioning, includes several of the five shortcomings discussed in this section of the paper. To provide an overview of the 26 studies, I have prepared Table 1, which lists each study and then denotes with an "X" the various shortcomings that I believe apply to a specific investigation. An appendix to the table lists the complete reference for each study, the type of study (prospective or cross-sectional), whether lead level was measured via blood or teeth, and the meta-analysis or meta-analyses in which the study was included. The five shortcomings listed in the body of the table correspond to the shortcomings that are discussed in the pages that follow. Again, I am not attempting to show that the studies with the most shortcomings found the largest IQ loss due to elevated BLL or that the studies with the fewest shortcomings found no relationship between BLL and IQ loss. I am just pointing out that the 26 "best" studies, regardless of outcome, tended to have a few important methodological shortcomings.

1.1. Shortcoming no. 1 — uncontrolled variables cloud conclusions drawn from even the best studies

Experimental designs have improved in controlling potential confounds in lead–IQ studies since the first set of studies began to appear decades ago, but there is still room for improvement. The variables that are undoubtedly most related to a young child's exposure to lead are the hardest to assess accurately. Such variables involve parenting skills, parenting styles of child rearing, parental time spent with the child, the skills and styles of key caretakers other than the parents, and so forth. These variables have usually not been measured in the 26 studies included in the meta-analyses. When specific parent-related and child-rearing variables have been measured, the results have been interesting.

In the Port Pirie study, Baghurst et al. (1992) and McMichael et al. (1994) found that the longer babies were breast-fed, the higher their WISC-R IQs and the lower their BLLs; parents living together was associated with lower BLL in both studies and with higher IQ in the Baghurst et al. study. Hatzakis et al. (1989) showed a significant relationship between "alcoholic mother" and both IQ and BLL; between "parents' divorce" and IQ; and between nailbiting and thumbsucking and BLL. In Fulton et al.'s (1987) study, noteworthy correlations were obtained between IQ on the British Ability Scales and measures (derived from a parental interview) of parent–child communication (.24) and parental participation with child (.33). Lansdown, Yule, Urbanowics, and Hunter (1986) reported significant correlations ($P<.01$) between BLL and such variables (obtained from a parental interview) as thumbsucking, age at weaning, quarrels/bickering, and family

Table 1
Summary of shortcomings associated with lead–IQ studies

		Parental IQ				
Lead–IQ study[a]	Used a global measure of SES[b]	Mother: used poor measure or none at all[f]	Father: did not test[g]	Did not control for multiple comparisons[c]	Compared "extreme" lead groups[d]	Lack of quality control in measuring children's IQ[e]
Baghurst et al., 1992			X			
Dietrich et al., 1993			X		X	
Ernhart et al., 1989		X	X			
Cooney et al., 1991		X	X			X
Bellinger et al., 1992		X	X	X		
Hatzakis et al., 1989	X		X		X	
Fulton et al., 1987		X	X		X	
Winneke et al., 1990	X	X	X			
Silva et al., 1988	X	X	X			
Yule et al., 1981	X	X	X			
Lansdown et al., 1986	X					
Harvey et al., 1988			X			X
Wang et al., 1989	X	X	X			X
Ernhart et al., 1985	X	X	X			
Schroeder et al., 1985		X	X			X
Hawk et al., 1986		X	X			X
Winneke et al., 1985	X	X	X			
Ferguson et al., 1988		X	X			
Smith et al., 1983			X			
McMichael et al., 1994			X			
Fulton et al., 1989	X			X		X
Needleman et al., 1979		X	X		X	X
Winneke et al., 1983	X	X	X	X	X	
Bergomi et al., 1989	X	X	X	X		
Pocock et al., 1987			X			
Hansen et al., 1989	X	X	X	X		

[a] Prospective studies appear first followed by cross-sectional studies. Complete citations along with the type study and additional information are presented in the Appendix.

[b] The indicated studies used mainly a global measure of SES (e.g., father's occupation, mother's education) instead of supplementing the global measure with a more specific assessment such as the HOME inventory or a parent questionnaire that involved direct contact with a caretaker.

[c] Valid scientific research requires that investigators hypothesize results, gather data, and form conclusions. Simply gathering data, subjecting it to multiple comparisons, and essentially "hunting" for results is not quality research. The indicated studies engaged in multiple comparisons, selected certain results while ignoring others, and failed to use a control for the multiple comparisons.

[d] The indicated studies compared extreme groups (high blood lead vs. low blood lead), while ignoring the groups in between.

[e] The indicated studies did not indicate the level of training of the examiners who administered the IQ tests to the children.

[f] The indicated studies either failed to measure maternal IQ or inadequately assessed maternal IQ with only a measure of picture vocabulary (e.g., PPVT-R, Quick Test) or with a group-administered test.

[g] The indicated studies failed to systematically assess the father's IQ.

cleanliness. Thumbsucking and family cleanliness were also significant correlates with BLL in Smith, Delves, Lansdown, Clayton, and Graham's (1983) study, as were children's play space and parent's relationship to baby.

Although many of the significant relationships in the aforementioned studies had a small effect size, they represent provocative results that suggest the need to measure, and, when necessary, to control subtle socioeconomic variables in the lead–IQ studies. However, many of the 26 studies measured socioeconomic status (SES) globally (for example, parents' education, father's occupation, a combination of education and occupation). It is good science to control for SES, but a global control is not enough. One should attempt to control for specific SES variables concerning the specific subjects in a given study. A number of research teams have attempted to do just this by administering the Home Observation for Measurement of the Environment (HOME) Inventory (Caldwell & Bradley, 1984), which requires an observer to go to the child's home and observe the interaction between the parent and child, ask questions of the parent, and observe the environment (e.g., the number of books in the home).

The use of the HOME inventory is a good thing, but there are several problems with the procedure: (a) there is no way to check on the parent's veracity in responding to questions about parent–child interactions; (b) the HOME inventory is intended as a screening test, based on a relatively brief visit, and is not comprehensive; (c) normative data are inadequate, based on small samples that are not representative of the nation as a whole; (d) there is little or no evidence that the HOME data are consistent from observer to observer or from time to time; and (e) usually, the HOME observation takes place at the time of the research study, when the child is older, and may not generalize to the earlier ages when the child ingested lead as an infant or toddler. Furthermore, as Banks, Ferretti, and Shucard (1997) point out, there may be confounding between some HOME items and exposure to lead, e.g., questions about physical environment (including cleanliness) may relate directly to lead exposure via lead in household dust. The net result is that "many HOME items can directly reflect the presence of lead and its effects, rather than measuring independent sources of variation in children's development" (Banks et al., 1997, p. 253). It is true that some of the criticisms of the HOME do not detract from its use in the lead–IQ studies. For example, its inadequate, non-representative normative sample is irrelevant when the HOME scores are used in a regression model, so long as the authors do not attempt to apply the regression equations outside of the sample in the study. In summary, the studies that have used the HOME inventory have done a good job of controlling for some of the confounding variance due to differences in parenting and the child's home environment; these studies have not, however, controlled for all of the pertinent variance because of shortcomings built into the instrument.

In all, 14 of the 26 studies made an effort to obtain some type of specific information about parenting: Nine of the 26 teams of investigators administered the HOME inventory (Baghurst et al., 1992; Bellinger et al., 1992; Cooney, Bell, & Stavron, 1991; Dietrich, Berger, Succop, Hammond, & Bornschein, 1993; Ernhart et al., 1989; Fergusson, Fergusson, Horwood, & Kinzett, 1988; Hawk et al., 1986; McMichael et al., 1994; Schroeder, Hawk, Otto, Mushak, & Hicks, 1985), and researchers from an additional five studies interviewed the parent in an attempt to obtain information about parent–child interactions, parenting styles, and so forth (Fulton et al., 1987; Harvey et al., 1988; Needleman et al., 1979; Pocock, Ashby, & Smith, 1987; Smith et al., 1983). The other 12 studies (see Table 1) relied on global indexes of SES and made no attempt to measure the specific kinds of socioeconomic and interactional variables that are most likely to relate directly to a young child's lead intake.

In addition to parenting variables, there are other variables that are known or believed to affect IQ substantially, and that may relate to a child's BLL, which are rarely controlled in the lead–IQ studies. A good case in point is otitis media (ear infections), especially when infants or toddlers have several incidents of acute otitis media and when the infections are accompanied by effusion (a collection of fluid in the middle ear) (Baldwin, 1993; Webster, Bamford, Thyer, & Ayles, 1989). Otitis media during the first 2 years of life affects a child's hearing ability (Friel-Patti & Finitzo, 1990), which, in turn, can have a negative impact on a child's development of language during the most crucial period for language development (Katz, 1978). Multiple episodes of otitis media, as well as otitis media with effusion, during the early years, are sometimes shown to be related to perceptual, cognitive, and language deficits (McShane & Plas, 1984; Rach, Zeilhuis, & Van den Broek, 1988; Updike & Thornburg, 1992). This topic has produced numerous research investigations that vary in quality and, like the lead–IQ studies, have produced conflicting results. Some studies have shown significant relationships between otitis media and diminished auditory perception (Updike & Thornburg, 1992), language ability (Holm & Kunze, 1969; Teele, Klein, & Rosner, 1984), intelligence (Teele, Klein, Chase, Menyuk, & Rosner, 1990), and reading ability (Teele et al., 1990; Updike & Thornburg, 1992). In contrast, perhaps an equal number of investigations have found no significant differences in functioning when comparing children with or without otitis media (Black & Sonnenschein, 1993; Fischler, Todd, & Feldman, 1985; Harsten, Nettelbladt, Schalen, Kalm, & Prellner, 1993; Roberts, Burchinal, Davis, Collier, & Henderson, 1991). The studies cited are illustrative; for reviews, see Roberts et al. (1991) and Webster et al. (1989).

The most comprehensive and well-controlled study of the effects of otitis media in the literature is the Greater Boston Otitis Media Study (Teele et al., 1984, 1990), a prospective study that enrolled 2568 consecutive children from 1975 to 1977, all of whom were given thorough ear examinations before age 3 months and at each subsequent visit. Teele et al. (1984) administered a battery of speech and language tests to a cohort of 205 three-year-old children, including those who had at least three episodes of otitis media with effusion by age 2 years and those who had no more than one such episode; their primary variable of interest was the number of days spent with middle ear effusion (ranging from 0 to >500 days). After controlling for several key confounds, multiple regression analysis indicated a significant association, in the predicted direction, between the total number of days with middle ear effusion and performance on speech and language tests (Teele et al., 1984). In a subsequent study by the Greater Boston research team, 207 children were randomly selected from 498 children who completed seven full years of observation, and were tested at age 7 on the WISC-R and on speech, language, and achievement tests (Teele et al., 1990). After controlling for SES and other pertinent confounds, the results of multiple regression analyses indicated significant relationships with speech, language, cognitive, and achievement variables. These differences were also substantial in magnitude. For example, adjusted WISC-R Verbal, Performance, and Full Scale IQs averaged about 112–113 for children with less than 30 days of middle ear effusion during the first 3 years of life, 107–108 for 30–129 days, and

104–106 for 130+ days (Teele et al., 1990). Like the presumed negative impact of lead on IQ, the observed negative effect of otitis media with effusion on children's neuropsychological functioning was significant and substantial for illnesses before age 4 years, but was non-significant for ear diseases at ages 4–7 years (Teele et al., 1990).

Despite the generally conflicting results of the accumulated literature on the effects of otitis media on children's functioning, the compelling findings of the well-designed studies by Teele et al. suggest that this medical variable be considered carefully when designing studies that relate BLL to IQ. Yet, only two of the 26 studies specifically controlled for persistent otitis media or illnesses affecting sensory function (Ernhart et al., 1989; Hatzakis et al., 1989). In the Hatzakis study, the "illness" variable proved to be a significant confound in the lead–IQ relationship; in the Ernhart study, a "Medical Problems" score correlated significantly with IQ.

Furthermore, in view of the known strong relationships between a mother's substance abuse or poor nutrition during pregnancy and the child's language development, cognitive ability, and behaviors (e.g., Sonderegger, 1992; Van Baar, 1990), it is surprising how relatively few lead–IQ studies systematically adjusted for confounds associated with prenatal factors. Of the 26 studies, only eight either specifically controlled for pregnancy risk factors or otherwise controlled for maternal smoking in general (Baghurst et al., 1992; Dietrich et al., 1993; Ernhart et al., 1989; Fergusson et al., 1988; Hansen, Trillingsgaard, Beese, Lyngye, & Grandjean, 1989; McMichael et al., 1994; Wang, Xu, Thang, & Wang, 1989; Winneke et al., 1983); a few other studies (e.g., Bellinger et al., 1992; Needleman et al., 1979) covaried tangential variables associated with pregnancy, such as birth weight or length of infant's hospital stay, but not the risk factors, per se.

The failure of most researchers to control for variables such as children's medical problems or prenatal care presents a challenge to the validity of the results of lead–IQ studies. Maternal smoking has been shown to correlate significantly with umbilical cord lead level and maternal lead level (Ernhart et al., 1989) and with children's tooth-lead level at age 6 1/2 years (Smith et al., 1983). In the Port Pirie study, Baghurst et al. (1992) and McMichael et al. (1994) found that higher IQs and lower BLLs were associated with both parents being non-smokers. In the study by Hatzakis et al. (1989), the child's history of chronic illness was a significant correlate of BLL, whereas the following medical variables were significant correlates of Full Scale IQ: birth weight, length of child's hospital stay after birth, history of CNS disease, history of head trauma, and illness affecting sensory function. Hatzakis et al. used as covariates in their regression model those variables, health-related or otherwise, that correlated significantly with BLL, IQ, or both. Similarly, Ernhart et al. (1989) covaried Medical Problems, even though it correlated significantly with children's IQ, but not with maternal, cord, or children's BLL. Technically, a confound in a lead–IQ study correlates significantly with both the dependent variable of children's IQ and with the independent variable of BLL. However, it is important to control for variables such as otitis media during infancy that are believed or known to vary alongside an outcome variable (such as IQ), even if their relationship to BLL is non-significant or unknown. If such variables (often referred to as "concomitant" variables) are ignored in prospective or cross-sectional studies of BLL and IQ, how can one feel confident that it is BLL, and not an uncontrolled variable, that is primarily responsible for an observed IQ loss in lead-exposed children?

Even more of a threat to the validity of the lead–IQ studies, however, are variables associated with intelligence that are either unknown or unmeasurable. There is a great deal that we do not know regarding the relationship of environmental variables to intellectual development. A good illustration is the bulk of research that has shown that children and adults are scoring higher on IQ tests from one generation to the next at the steady rate of three points per decade within the United States (the rate is higher — sometimes twice as high — in many other developed nations in the world; see Flynn, 1987; Kaufman, 1990, Chapter 2). When the research was first published by Flynn (1984), the data were based on individuals tested between the 1930s and 1970s. The explanation that was widely given to what has now become known as the "Flynn effect" seemed simple enough: In the 1930s and 1940s, there was far less environmental stimulation than in the 1960s and 1970s owing to technology (television and other forms of mass media) and changes in parenting styles (increased awareness of the importance of providing cognitive stimulation in infancy). However, as the research has continued to accumulate, it has become increasingly clear that the 3 points per decade gain did not stop in the 1970s. Rather, it has continued unabated, and was just as constant a gain from 1987 to 1997 as it was from 1947 to 1957. Quite clearly, the explanation cannot simply be greater exposure to mass media. The precise variables that are responsible for the steady gain in human intelligence from generation to generation are unknown. Some theorists (Lynn, 1990) have speculated that improved nutrition is largely responsible, and others (e.g., Flynn, 1987) offer alternate hypotheses, but current research has yet to answer the question empirically. Yet, one thing is clear: The explanations involve environmental variables because differences as large as 3 points per decade in the US, and 6 to 8 points per decade for five European countries (Kaufman, 1990, Fig. 2.2), are far too large to be explained by genetic or evolutionary variables.

Why have I dwelled so much on an issue concerning improved IQ when the lead research suggests loss of IQ? For a few main reasons: First, whatever environmental variables are responsible for the steady gain are likely to lead to loss of intellectual functioning when they are diminished or withdrawn. Next, the generational research demonstrates that sometimes changes of a few points in IQ occur in the absence of a known cause; yet, even though people were so sure that IQ gain was due to the introduction of television and the escalation of mass media, hypotheses are sometimes wrong. In this instance, the gain per decade of 3 points is similar in magnitude to the size of the effect attributed to BLL.

Another consequence of the research findings on generational shifts in IQs concerns lead studies that used maternal IQ as a variable in the study, either as a covariate or as a "target IQ" for comparison with their lead-exposed child's IQ. An example of the latter kind of study is an investigation by Bellinger and Needleman (1983), one that was not included in the meta-analyses but that was, nonetheless, conducted by one of the leading research teams on the possible adverse effects of lead. They used IQ based on the Peabody Picture Vocabulary Test (PPVT; Dunn, 1965) as the estimate of maternal IQ and the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) to measure the children's IQ. Then they considered the maternal IQ as the IQ that is "expected" for the child. This approach is poor for several reasons, one of which is discussed in the next section: The PPVT is not even an IQ test. However, another reason concerns the fact that cognitive abilities change over time. The PPVT was normed in 1959. When the data were gathered for the Bellinger and Needleman study, in the mid-1970s, the PPVT norms were already more than 15 years old, and about 5 points out of date — which means that they would overestimate the mother's PPVT "IQs" by about 5 points. In contrast, the WISC-R had just been normed in 1972, so the test was at most out of date by 1 point. Therefore, the mother's IQs should have been adjusted by 4 points to be comparable to the WISC-R IQs of their children (i.e., subtract the 1-point out-of-datedness of the children's scores from the 5-point out-of-datedness of the mother's scores). Bellinger and Needleman found that the children with elevated lead levels had IQs that were 3.94 points lower than expected based on maternal IQs. That is precisely the amount of points that should have been corrected because of the different years in which the PPVT and WISC-R were normed. The entire "significant" difference is nothing but an artifact of the methodology. (Although mean maternal IQs are not provided by Bellinger and Needleman, 1983, examination of their scatter plot and data presented by Needleman et al., 1979, for an overlapping sample, indicate an average of about 110 on the PPVT. Had the mean been <100, then the effects of regression to the mean would have had to be accounted for when interpreting whether or not lead level was associated with decreased IQ.) Similar problems involving failure to control for the Flynn effect have led to incorrect conclusions regarding other popular topics, e.g., overestimating the superiority in intellectual ability of Japanese vs. American individuals (Flynn, 1983; Lynn, 1982).

Ultimately, though, the point goes beyond any one topic I have raised here. My overall contention is that most studies of lead and IQ, even when limited to the best available studies, have failed to control for important parenting variables, subtle socioeconomic variables, and medical variables; also, they have been unable to control for what are undoubtedly a plethora of unknown but potentially potent variables. The net result is to interpret the results of the lead–IQ studies with more caution. This point is made even clearer by understanding one significant limitation of the multiple regression procedures that have been used in nearly all of the 26 studies that have been included in the meta-analyses. The investigators have uniformly used appropriate, sophisticated regression analysis procedures. They have identified various sets of potentially confounding or contaminating variables and have entered these variables into the regression equation in an attempt to control as much of the unwanted variability as possible. Then, lead level is added to the equation to see if it contributes a significant amount of variance to the prediction of IQ, over and above the prediction that is obtainable from the confounds alone. When lead level adds significantly to the equation, researchers conclude that lead level leads to IQ loss. However, this conclusion is not entirely accurate.

A more correct statement is that the significant increase in prediction is due not only to lead level, but also to all other variables, known or unknown, not controlled in the study. As the history of lead–IQ research has revealed, the more that pertinent confounding and contaminating variables are identified and controlled, the smaller the relationship between lead level and IQ loss; it is a common finding for significant relationships between BLL and IQ to become much smaller (e.g., Baghurst et al., 1992), or to disappear altogether (e.g., Ernhart et al., 1989), when controlling for a diversity of relevant variables. Yet, so many variables that are usually uncontrolled in the lead research bear obvious potential relationship to the ingestion of lead, such as parental supervision of infants and toddlers. Furthermore, so many environmental variables associated with IQ are unknown — witness the increase in IQ across generations. It is possible that some or all of the IQ loss due to BLL is due to these other uncontrolled variables.

Low BLLs have sometimes been interpreted as causing IQ loss in children (e.g., Schwartz, 1994). The issue of causality is complex and difficult either to prove or disprove. The existence of uncontrolled or poorly controlled variables, even in the best lead–IQ studies, complicates the interpretation of lead as a causal agent in decreasing children's' IQs. As Reynolds (1999) points out concerning the related topic of the possible harmful effects of maternal smoking on the developing infant, "An unknown or poorly assessed variable may be associated with the cause of both variable sets giving rise to a mirage of causality when in fact it resides elsewhere." Within the lead–IQ literature, two of the meta-analysis studies present opposing interpretations of the literature regarding causality. Schwartz (1994) states: "Studies in primates have also documented disturbances in cognitive functioning at relatively low blood lead levels . . . these are experimental protocols, with no confounding by omitted variables to worry about. In the light of this evidence, it is only reasonable to assume the strong epidemiological association found in the meta-analysis is causal" (p. 53). Pocock et al. (1994) reach a different conclusion: "Observational epidemiology cannot distinguish between this direction of effect and the more important issue, 'does lead cause a deficit in IQ?' However, this review provides some implicit evidence that reverse causality is plausible" (p. 1196). Resolving the causality issue is beyond the scope of this paper.

1.2. Shortcoming no. 2 — parental IQ is typically measured poorly or not at all

Lead researchers have become aware that one of the strongest correlates both of IQ and lead level in a young child is that child's parents' IQs, and that this potential confound must be controlled in lead–IQ studies. Parents' IQ is related to SES and to genetic factors; controlling for it as a confound, even when SES is otherwise controlled, is important for a good research design. Although most lead researchers realized this desirability, it is nonetheless true that eight of the 26 studies failed to measure parental IQ (Bergomi et al., 1989; Fergusson et al., 1988; Hansen et al., 1989; Wang et al., 1989; Winneke, Brockhaus, Ewers, Kramer, & Neuf, 1990; Winneke et al., 1983, 1985; Yule, Lansdown, Millar, & Urbanowicz, 1981). Furthermore, despite the rigorous attempts that most experimenters made to use state-of-the-art IQ tests for the children in the studies, a similar rigor was not followed when assessing their parents. Only two research teams administered the accepted criterion of adult intelligence — the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981), which has since been superseded by the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997) — and both of these teams were reporting on the Port Pirie study (Baghurst et al., 1992; McMichael et al., 1994).

An additional five studies used a two-subtest short form (Vocabulary and Block Design) of the WAIS (Wechsler, 1955) or WAIS-R as the estimate of parental intelligence (Dietrich et al., 1993; Hatzakis et al., 1987; Lansdown et al., 1986; Pocock et al., 1987; Smith et al., 1983). In view of the fact that the WAIS and WAIS-R include 11 subtests, the elimination of nine subtests forces the researchers to estimate the parents' IQs from only a small portion of the complete battery. Even a good short form such as the one used in these studies makes substantial errors when estimating the Full Scale IQ that would have been obtained had the complete battery of 11 subtests been administered. Standard errors of estimate for the Vocabulary and Block Design short form average about 6 points, which means that the parents' obtained IQ on the short form has a two out of three chance to be within 6 points (in either direction) of the IQ the parent would have earned on the complete test battery. Nonetheless, it is true that both Vocabulary and Block Design are both excellent measures of "g" or general intelligence, and, taken together, they correlate above .90 with Full Scale IQ (Kaufman, 1990). The substitution of a Wechsler short form for the full battery, though not ideal testing practice, is still an acceptable procedure for use in multiple regression analysis and will not adversely affect the outcome of the study.

However, the acceptability of using a brief version of a Wechsler scale does not generalize to most other brief measures used to measure parents' intelligence in a number of lead–IQ studies, most notably the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981). This is a one-subtest measure that is not an intelligence test. Dunn and Dunn (1981) state in the test manual, "The PPVT-R is designed primarily to measure a subject's receptive (hearing) vocabulary ... It is not, however, a comprehensive test of general intelligence" (p. 2). It yields "standard score equivalents," not IQs, and is misused as a measure of general intelligence. In addition, the PPVT-R has two other shortcomings: (a) each item is a four-option multiple-choice question, which allows chance guessing to play a potentially large role in one's obtained score; and (b) despite possessing an excellent and representative normative sample for children, its norms for adults are poor — the samples of adults are relatively small and non-representative of the nation on the variables of geographic region and ethnic group, and virtually all of the adult sample was tested in a group format (even though the test is intended to be given individually) (Kaufman, 1990, pp. 606–607). Nevertheless, this quick-and-easy test (or its predecessor, the PPVT) was used as the sole measure of parents' intelligence in five studies (Bellinger et al., 1992; Cooney et al., 1991; Ernhart et al., 1989; Hawk et al., 1986; Needleman et al., 1979); the similar Quick Test (Ammons & Ammons, 1962) was used in two studies (Ernhart, Landa, & Wolf, 1985; Schroeder et al.,1985); and a group-administered test of verbal ability was used in one study (Silva, Hughes, Williams, & Faed, 1988).

One might argue that picture vocabulary tests, even if not intended to measure IQ, correlate highly enough with conventional IQ tests to be an adequate substitute for an IQ test. Typically, global IQs on different IQ tests correlate in the .70 to .80 range (e.g., Sattler, 1988, Table 6-3). The median value for the PPVT-R is in that ballpark (.68 in Sattler's table), but the range of correlations for the PPVT-R is unusually large from study to study (Robertson & Eisenberg, 1981, Table 4.4). It is not uncommon for coefficients to dip below .40 in PPVT-R studies, and sometimes below .20. In addition, coefficients have tended to be lowest for samples of economically disadvantaged individuals, the type of sample that is typically included in lead–IQ studies. Based on 68 coefficients between the PPVT-R and a variety of IQ tests for economically disadvantaged samples, values ranged from .13 to .87 with a median of .51 (Robertson & Eisenberg, 1981, Table 4.4). These data suggest that the PPVT-R or similar tests that do only a fair job of predicting Wechsler's Full Scale IQ are not adequate substitutes for conventional intelligence tests for measuring parents' intelligence in lead–IQ studies, especially since effect sizes of 0.15 to 0.25 S.D. (2 to 4 points) are anticipated. If the effect sizes were larger, say 0.5 S.D. or more, then the .50–.60 correlations with IQ would not be a major issue. However, with small effect sizes disproportionality of the error variance in the measurement of a control or confounding variable and the dependent variable can lead to artifactual conclusions, artifacts of the disproportionality of the two error variances. In essence, one is given the facade of statistical control when it is, in fact, absent or more severely limited than it appears superficially.

Overall, parents' IQ was not assessed in eight studies and was assessed inadequately (i.e., by a picture vocabulary test or a group-administered test) in nine more studies. These 17 studies are noted in Table 1 as having shortcomings. The studies that used brief measures that are known to correlate highly with IQ (e.g., matrices and vocabulary tests), like the studies that used WAIS or WAIS-R short forms, are considered to have measured parents' intelligence adequately.

In addition, despite the desirability of controlling for both parents' IQs in lead studies, only a single study systematically tested both fathers and mothers (Lansdown et al., 1986, who administered the WAIS short form to both parents). With the exception of a few studies that tested the caretaker of the child (typically the mother), virtually all attempts to control for the potential confounding variable of parents' IQ involved the assessment of maternal IQ only. Yet, data from genetic investigations indicate that both parents are needed to get the best estimate of the genetic contribution to the child's intelligence: When children and parents live together, children's IQs correlate .42 with one parent's IQ vs. .50 with the average of both parents (Kaufman, 1990, Table 2.1). The difference between .42 and .50 is noteworthy. The former value denotes that the parent's IQ accounts for 17.6% of the variation in the child's IQ, whereas the latter value denotes that both parents' IQs account for 25.0% of the variation in their child's IQ. Testing the IQ of the father, as well as the mother, therefore, represents a 42% increase in the overlap between parent and child IQ. Furthermore, when correlation coefficients are translated into effect size, a value of .42 corresponds to a medium effect size; a coefficient of .50 corresponds to a large effect size (Cohen, 1992).

It is true that there are practical issues involved in getting cooperation from parents for research studies (see Ernhart et al., 1989), but that does not excuse using a test that is not intended as an IQ test (PPVT-R), or virtually eliminating fathers from consideration. The net conclusion is that even though researchers have claimed to control for the key potentially confounding variable of parents' IQ, a majority of the best studies have done an unimpressive job of it. Parental IQ is not only substantially correlated with children's IQ, it also correlates significantly with BLL (e.g., McMichael et al., 1994), making it an essential covariate in lead–IQ studies.

1.3. Shortcoming no. 3 — failure to control for multiple comparisons

Competent experimental research requires that the investigators hypothesize the results they are anticipating before they examine the data, not after they examine the results of the study. It is simply bad science to conduct "multiple analyses" and then interpret only the ones that give the answers the researchers were seeking. The type of research that involves conducting many analyses at once, and then picking and choosing the analyses the experimenters like best, is known informally as a "shotgun approach." It is appropriate to use this type of approach when conducting exploratory research, as a means of generating hypotheses for future investigation, not as a means of drawing causal conclusions. Yet, even when a shotgun-type approach is used appropriately, researchers need to be cautious when interpreting apparently significant findings in the data. Whenever many analyses are conducted simultaneously, the odds of finding one with significant results increases dramatically, simply because the shotgun approach allows experimenters to take advantage of chance error. The more analyses conducted, the greater the likelihood of finding a significant relationship. Error variance cannot be related meaningfully to an outcome, but error variances are sometimes correlated purely by chance. The more analyses done, the more likely correlated error variances are to occur. Because it is considered unscientific to take advantage of chance error, another term for correlated error variances, correct research methodology requires experimenters to control for the chance errors that necessarily creep in when conducting simultaneous multiple comparisons. One popular such correction (albeit a conservative one) is known as the Bonferroni procedure, although other techniques are applied as well (Lehman, 1991).

Unfortunately, five of the 26 lead–IQ studies made this precise kind of "multiple comparisons" error (see Table 1): The researchers used a shotgun approach, and then honed in on the specific significant results that were consistent with their a priori hypotheses about the importance of lead level on a child's intellectual functioning; the researchers in these studies made no attempt to control for the multiple comparisons. The Bellinger et al. (1992) study serves as a good illustration of this type of error in experimental design. These investigators related blood level at seven points in time (cord, 6 months, 12 months, 18 months, 24 months, 57 months, and 10 years) to WISC-R Verbal, Performance, and Full Scale IQ (Bellinger et al., 1992, Table 2). Thus, they made 21 simultaneous comparisons of which only two were statistically significant (WISC-R Verbal and Full Scale IQ at age 24 months). When the obtained probabilities are corrected by the Bonferroni procedure (which the authors did not do), neither of the two WISC-R relationships in their Table 2 remain significant at the .05 level. When applying the Bonferroni procedure, one divides the family-wise alpha level that is desired (let us say .05) by the number of simultaneous comparisons (in this case, 21). To achieve a family-wise alpha level of .05, Bellinger et al. would have had to obtain $P < .0024$ (i.e., .05 divided by 21) for any single comparison. Neither of their two significant comparisons qualified. The authors likewise made 28 multiple comparisons with achievement scores from the Kaufman Test of Educational Achievement (K-TEA; Kaufman & Kaufman, 1985), reported in Table 3 of their article, but again failed to correct for the chance error. In this instance, they obtained three significant findings, but if the Bonferroni correction is applied (requiring $P < .0018$ for each comparison to yield a family-wise error rate of .05), only two differences remain significant. It is true that Bellinger et al. offer possible explanations for the significant results at only one of the seven ages in the Discussion section of their article (although they never mention the obvious multiple-comparison/chance-error explanation), and they advise that the significant findings "should be interpreted cautiously until confirmed by other studies" (p. 859). However, the authors do not heed their own advice. In the Abstract to the article, they feature the significant findings at age 24 months without once advising caution, and little caution has been followed when the results at age 24 months have been the only results interpreted in the meta-analyses.

The failure to control for multiple comparisons is a serious methodological shortcoming in the lead–IQ research. This type of mistake occurred in a few very well-controlled investigations conducted subsequent to the meta-analyses (e.g., Tong et al., 1996; Wasserman et al., 1998); hopefully, it will be rectified in future investigations.

1.4. Shortcoming no. 4 — comparison of the IQs of the two extreme "lead-level" groups

An error in experimental procedure that is similar to the multiple comparison procedure described in the preceding section concerns the comparison of mean IQs earned by those children who have the lowest BLLs with the mean IQs of those children who have the highest BLLs. Again, such comparisons are not warranted when multiple regression methodology is applied. The regression procedure involves relationships among predictor variables (lead level + the potential confounds) and the criterion variable (typically IQ) for the total sample of children in the study. This technique permits interpretation of data for the total sample. It does not allow the experimenter to pluck out a portion of the sample — such as the two groups with extreme lead levels — and eliminate the rest (usually the bulk) of the sample.

Even though the IQs for the extreme groups have been adjusted for the potential confounding variables, those adjustments were based on the large total sample. In order to properly compare the two extreme groups (which, together, now comprise a new "total" sample that is considerably smaller than the original one), a new set of adjustments must be made for the confounding variables. Otherwise, there is no way of knowing that the "high lead" and "low lead" groups are properly matched on SES, maternal IQ, and so forth. Without the assurance that the two extreme groups are matched with each other on the confounding variables, the most probable explanation of a mean IQ difference between them is one or more of the confounding variables. Unfortunately, none of the researchers who have compared the mean IQs of the extreme groups have conducted the pertinent additional multiple regression analyses based only on the two extreme groups (see Table 1 for a list of the five investigations that included this shortcoming).

As with the problem of multiple comparisons, the effect of chance errors looms very large when only data from the extreme groups are analyzed to determine the effect of lead level on IQ. It is possible to compare groups who differ in lead level based on the reanalysis of data, by making new adjustments for confounds, but even better as a technique for group comparison is analysis of variance (ANOVA) methodology, as employed in the study conducted by Smith et al. (1983). In that study, three groups were specified in advance (total $N=403$), they were matched statistically on pertinent variables, the relevant statistical procedure (multiple analysis of variance and covariance) was applied, the comparisons among pairs of groups were specified in advance, and corrections for simultaneous multiple comparisons were built into the procedure. Smith et al. observed significant differences of about 4–5 points on the three WISC-R IQs, in favor of the Low BLL group relative to the High BLL group (with the Medium BLL group scoring about midway between the two), prior to covarying for confounding and concomitant variables. After controlling for a diversity of variables, the High and Medium BLL groups earned nearly identical mean IQs, while the Low BLL sample scored a non-significant 1–3 points higher.

In contrast to the Smith et al. (1983) study, several teams of investigators used multiple regression methodology inappropriately when they compared the two extreme "lead-level" groups, as indicated in Table 1. Fulton et al. (1987) actually divided their sample into 10 subgroups and then compared the two extremes. Winneke et al. (1983) compared three BLL groups (low, middle, and high in terms of tooth leads) on WISC-R Verbal IQ, and stated, "after correction for confounding, these differences [were not] significant ($F=0.91$; $P=.41$)" (p. 240). Despite the non-significant differences among the three subsamples, Winneke et al. included in their Abstract a comparison of the two extreme tooth-lead groups, noting the "tendency for high-level children . . . to be inferior to low-level children . . . by 4.6 IQ points after correction for confounding" (p. 232). However, the correction was for all three groups, not the two extremes, and the obtained F value was quite low, not suggestive of any tendency. Whenever mean differences for extreme groups are obtained from multiple regression studies in the manner described here, such mean differences are uninterpretable and cannot be used meaningfully to reach inferences on the relationship between lead level and IQ.

1.5. Shortcoming no. 5 — lack of quality control in measuring children's IQ

Conventional clinical IQ tests that are administered individually to referred children and adults, such as Wechsler's scales, are not like most other psychometric measures used in research investigations. They cannot be given by untrained examiners or even by examiners who are provided with specific in-depth training prior to the data collection. As stated in the WISC-III manual (Wechsler, 1991): "Because of the complexities of test administration, diagnosis, and assessment, examiners who use the WISC-III should have training and experience in the administration and interpretation of standardized, clinical instruments, such as the WISC-R or other Wechsler intelligence scales. They should also have experience in testing children whose ages, linguistic backgrounds, and clinical, cultural, or educational histories are similar to those of the children they will be testing . . . Furthermore, examiners should be familiar with the *Standards for Educational and Psychological Testing* (American Psychological Association, 1985)" (p. 10). Valid administration and interpretation of the WISC-R, WAIS-R, WISC-III, and other clinical IQ tests requires professional, graduate-level training in psychometrics, clinical skills, child development, neuropsychology, and so forth. This level of training is not essential for all standardized tests. For example, the K-TEA, a test of educational achievement administered in the Bellinger et al. (1992) study, does not require the same level of training, and does not carry the same restrictions regarding its purchase and use, as does an individually administered intelligence test, such as the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983), the British Ability Scales (Elliott, Murray, & Pearson, 1979), or a Wechsler scale. However, for an IQ test such as the WISC-R or WAIS-R, the validity of the data is in large part a function of the person who administers and scores the test. Furthermore, even experienced examiners are prone to making clerical errors when scoring the tests due to carelessness (a pervasive, serious problem; see Kaufman, 1990, pp. 101–105; Kaufman, 1994, pp. 127–128).

One potential validity problem that might result from using inexperienced examiners was noted in an early lead–IQ study conducted by Gregory, Lehman, and Mohan (1976). These investigators, by happenstance, discovered that the examiner's attitude and demeanor had a greater impact on IQ than the lead level in the child's blood. They noticed that one of the five graduate student examiners came up with relatively low scores for the children tested (average IQ of 90), and another consistently produced inflated scores (average IQ of 104). The first examiner "was very formal, precise, cold, and hurried," while the second offered "support and encouragement that bordered on leading the subjects to the correct answer" (Gregory, 1987, p. 154).

Unfortunately, from my first-hand knowledge of large-scale research projects, it is not uncommon for inexperienced examiners (or, at least those without proper, in-depth training) to be used to collect IQ data in research studies. That situation makes it feasible that examiner errors such as the one just described (even if not as extreme) occurred in some of the 26 "best" lead–IQ studies. It is important for any study that uses an individually administered IQ test such as Wechsler's scales as an independent or dependent variable to indicate the qualifications of the examiners and to institute some type of quality control to ensure the accuracy of the data. Yet, the quality of the IQ test administrations and the concomitant validity of the obtained IQs is rarely assessed thoroughly or systematically. Winneke et al. (1990) provided a thorough and impressive set of quality assurance procedures "to improve the comparability of the psychological test data" (p. 555), thereby providing a notable — and important — contrast to the lack of attention given to this crucial validity variable by most other lead–IQ research teams. Indeed, in two previous investigations by Winneke et al. (Winneke et al., 1983, 1985), advanced psychology students did the testing with no indication that they were given appropriate supervision or that quality control procedures were applied. Apart from the Winneke et al. (1990) study, quality control of the IQ data was directly addressed in only four other studies: (a) Hatzakis et al. (1989) mentioned that "Most of the neurobehavioural tests administered, and the quality control procedures, conformed to the WHO/UNDP Protocol (WHO, 1984)" (p. 214), a protocol that also was instrumental in identifying the extensive quality assurance procedures instituted by Winneke et al.; (b) Fulton et al. (1987) stated that, "Tests were done in school by two trained psychologists whose performance was checked in a crossover trial" (p. 1222); (c) Smith et al. (1983) indicated, "A reliability trial between testers was carried out on 60 children from two schools. Testers were shown to be reliable" (p. 7); and (d) Bergomi et al. (1989), adhering to the WHO (1984) protocol, trained the four psychologists "before the study to avoid inter-tester variability ... In addition, the psychometric tests which are most sensitive to subjective evaluation ... were scored by all four psychologists and the scoring was supervised by a psychologist in the coordinating group within the WHO project" (p. 183).

Authors of several studies showed awareness of the need for well-trained or experienced IQ examiners, even if no additional mention was made of quality control; these included Ernhart et al., (1985, 1989), Fergusson et al. (1988), Hansen et al. (1989), Lansdown et al. (1986), McMichael et al. (1994), Silva et al. (1988), and Yule et al. (1981). A few other research teams simply mentioned that a psychologist or psychometrician did the testing, without specifying the person's level of training (Baghurst et al., 1992; Bellinger et al., 1992; Dietrich et al., 1993; Pocock et al., 1987). All of these studies are given credit for, at the least, showing awareness that administering IQ tests requires specialized training.

Table 1 lists the seven studies that failed to demonstrate any type of awareness that individually administered IQ tests demand extra care to obtain valid data. Authors of these studies made no mention of who administered the tests, even though pediatricians, nurses, and other health personnel were frequently named for conducting different aspects of medical examinations. Harvey et al. (1988) videotaped all cognitive testing, a procedure that was not used during the standardization process, thereby violating normative procedures and possibly invalidating the IQ data. The list of studies in Table 1 considered to have fallen short in quality control is a conservative list. The two studies by Winneke et al. in which students were used to gather the data are not included in Table 1 because they were advanced psychology students. Further, merely noting that the examiner was a "psychologist" (Bellinger et al., 1992; Pocock et al., 1987), an "experienced psychometrician" (Dietrich et al., 1993), or a "research psychologist" (Baghurst et al., 1992) neither suggests nor implies that the examiner received appropriate supervised graduate-level clinical training in assessment or that appropriate care was taken to ensure accurate scoring.

In addition, several studies are excluded from Table 1 even though the investigators opted to administer short forms of Wechsler's children's scales instead of the whole battery. Three studies administered a four-subtest short form of the WISC (Winneke et al., 1985, 1990) or WISC-R (Bergomi et al., 1989), and a fourth study administered an eight-subtest WISC-R short form, eliminating two subtests — Picture Arrangement and Comprehension (Silva et al., 1988). For the four-subtest short forms, the standard error of estimate for predicting the IQ the child would have earned on the complete test battery is ± 5 points. And even though the error of estimate is smaller for the eight-subtest short form (about ± 3 points), the two particular subtests eliminated are the two best measures of real-life, common-sense social intelligence included in the Wechsler scales (the precise kind of tasks that one would most want to include if the goal is to translate IQ scores to future functioning in society). The use of short forms does not affect the validity of the obtained results in the lead–IQ studies because all of the short forms used correlate more than .90 with Full Scale IQ, a point mentioned previously regarding the use of short forms to measure parents' IQ. However, the complete Wechsler Full Scale is the accepted criterion of children's intelligence, not a truncated version. Eliminating subtests as a time-saving device is understandable in a research study, but it is, nonetheless, ill advised for the main outcome variable, especially when the results of the studies are used to set public policy.

In fact, the lack of quality control in some, perhaps most, of the best lead–IQ studies actually decreases the probability of finding a significant effect because anything that introduces random as opposed to systematic error reduces the odds of finding an effect. Within the lead–IQ research, this effect is well known. Greene and Ernhart (1993) note "the downward bias in the magnitude of the estimated lead effect due to measurement error" (p. 405); further, there is evidence that significant lead effects remained after accounting for measurement error (Atkinson, Crocker, & Needleman, 1987). The "null-biasing" effects of random error are not the issue. As I stated earlier, I have identified shortcomings in the 26 best lead–IQ studies, without regard to the positive or negative findings of each study regarding the impact of BLL on IQ loss. My concerns are: (a) that the same care has not been taken in the measurement of the main psychological variable, the outcome variable of IQ, as has been taken in the measurement of the main medical variable, the independent variable of lead level; and (b) the possible introduction of random error into the regression analyses because of a lack of appreciation of the subtleties of the administration and scoring of clinical tests is bad science.

1.6. Overview

The preceding five points, taken together, represent a rationale for interpreting all of the lead–IQ research, both the positive and the negative findings regarding the impact of low BLLs on IQ, with caution. Table 1 includes six columns to depict the five shortcomings (measurement of parental IQ separates the adequacy of measuring maternal IQ from whether or not fathers were tested). Of the 26 studies, all made at least one of the six possible errors and 17 displayed three to five of the shortcomings. The following five studies demonstrated four or five shortcomings (Bergomi et al., 1989; Hansen et al., 1989; Needleman et al., 1979; Wang et al., 1989; Winneke et al., 1983).

If studies are divided by whether they yielded positive or negative outcomes, it is clear that both types of studies displayed shortcomings. Of the 26 studies, 14 showed significant IQ loss after controlling for confounds. Three studies gave equivocal results: Winneke et al. (1990) found only borderline associations between BLL and IQ loss; Schroeder et al. (1985) found a significant lead–IQ association in their initial study but not in a 5-year follow-up; and Fergusson et al. (1988) found significant associations between BLL and achievement but not with IQ. Nine studies produced negative results regarding the relationship of BLL to IQ: Cooney et al. (1991), Ernhart et al. (1985, 1989), Harvey et al. (1988), Lansdown et al. (1986), Pocock et al. (1987), Smith et al. (1983), and Winneke et al. (1983, 1985). Of the 17 studies with three to five shortcomings, 11 gave positive results, four gave negative results, and two gave equivocal results. Of the nine studies with the fewest shortcomings (one to two), three gave positive results, five gave negative results, and one gave equivocal results.

However, these simple numbers do not tell the whole story and are meant to provide an illustrative overview. For example, failure to control for important variables — known or unknown — affects every study in Table 1, but the only ones marked with an "X" are the studies that relied solely on a global measure of SES instead of obtaining more specific or subtle data.

One might look carefully at the study that probably controlled more important SES and family-related variables than any other lead–IQ investigation: Bellinger et al. (1992). In addition to the HOME inventory, that team of investigators administered to parents the Family Adaptability and Cohesive Evaluation Scale (FACES; Olson, Portner, & Lavee, 1985), Social Readjustment Rating Scale (Holmes & Raye, 1967), Parenting Stress Index (Abidin, 1986), Children's Life-Events Inventory-Revised (Chandler, 1982), and Social Support Network Inventory (Flaherty, Gaviria, & Pathak, 1983). All of these measures help identify the relationship of many potential child-related and parent-related variables to the child's IQ. In that regard, these experimenters went well beyond the norm established in lead–IQ studies for the control of key confounding variables, and did an excellent job of controlling for variables that have been largely ignored by other lead researchers.

Yet, despite these bold efforts, the study has some noteworthy pitfalls. As noted previously, Bellinger et al. (1992) assessed parents' intelligence with a test that was not intended as an IQ measure, they tested only the mother, and they apparently used the long-outdated, original PPVT (Dunn, 1965). In addition, they did not correct for the many statistical comparisons they made simultaneously. However, there are other problems as well. The authors used the WISC-R, which ordinarily would be a good choice of IQ test in view of its stature and validity, but the data were collected around 1990–1991, about 18 years after the test's 1972 standardization. Because of the "Flynn effect," in 1990, the WISC-R was yielding IQs that averaged about 5 points too high for American children; i.e., the norms were 5 points out of date. The authors would have been wise to consider using a newer measure of intelligence such as the Woodcock–Johnson-Revised Tests of Cognitive Ability (Woodcock & Johnson, 1989). At the least, they should have incorporated the spuriously high WISC-R IQs into their interpretation of the data.

In addition, the study included children (95% of whom were white) from relatively higher SES environments. Their average IQs of 116–119 reflect "High Average" functioning, even when the IQs are corrected for the 5-point spuriousness due to outdated norms. The technique of multiple regression is very sensitive to the specific sample tested, and the results for one sample are not readily generalizable to samples that differ in meaningful ways from the original sample. Therefore, the results of the Bellinger et al. (1992) study only generalize to similar samples of high SES white children, not to minorities or to whites from lower social classes. Generalizations from multiple regression studies conducted in other countries are likewise not generalizable to children living in the US. In view of the fact that the meta-analyses include studies conducted in countries such as Greece, Germany, Great Britain, Denmark, China, Australia, Scotland, and New Zealand, any generalizations to American children from the merged data sets are tenuous at best. For example, the relative contributions of genetics and environment to a child's IQ differ, sometimes markedly, from country to country. The genetic component tends to be population specific and may be larger in countries that are homogeneous in their population (like Denmark or Norway) than in the US (Kaufman, 1990, Table 2.2). If the average contribution of environment to children's IQ is different in Europe than in the US, then the regression coefficients obtained in European studies for socioeconomic and parenting confounding variables are not generalizable to American children.

The criticisms of the Bellinger et al. (1992) study are not intended to demean a research team that has made prodigious contributions to the understanding of the possible relationships of lead to children's cognitive and behavioral functioning (Bellinger, Leviton, Needleman, Waternaux, & Rabinowitz, 1986; Bellinger, Leviton, & Waternaux, 1989; Bellinger, Leviton, Waternaux, Needleman, & Rabinowitz, 1987; Bellinger, Needleman, Bromfield, & Mintz, 1984; Bellinger et al., 1991; Needleman, 1982, 1987; Needleman, Geiger, & Frank, 1983; Needleman et al., 1996). Rather, the cautions are intended to underscore that even the best studies on the topic need to be interpreted within the context of the limitations of empirical research and instrumentation and the need to strive continually for improved methodology, especially when the results of such research are applied directly to public policy.

2. Questionable interpretations of the IQ loss attributed to low lead levels

Despite the research shortcomings inherent in the diversity of lead–IQ investigations, the results of these studies have been instrumental in setting public policy. Included in public documents are statements that imply a linearity in the relationship between blood-lead concentration and IQ loss, that encourage fragmentation of a single IQ point into its component parts, and that express concern about the societal impact of small IQ losses. Consider the following two quotes from a recent set of guidelines proposed by the

Environmental Protection Agency (EPA) on the topic of "Lead; Identification of Dangerous Levels of Lead" (Federal Register, 1998): "EPA assigns risk reduction value to fractional losses of an IQ point — tenths and even hundredths of a point" (p. 30320); "The computation of IQ point loss is based on an average decrease of 0.257 IQ points per increase of one μg/dl in blood-lead concentration" (p. 30321). Or, from an earlier document, "preventing a one μg/dl increase in a 1 year old child's blood lead level saves $1493 ... in lifetime earnings" (Federal Register, 1996, p. 29103). In addition, as noted earlier in this paper, lead researchers have expressed concern about the potential deleterious societal impact of IQ loss due to exposure to low levels of lead in the environment, leading to enormous increases in low-IQ children and decreases in high-IQ children (Needleman, 1989).

2.1. There is no documented linear relationship between lead level and IQ

Because of the realities of how the results of the lead–IQ studies have been used to help set public policy, I will proceed as if the results of the studies are valid, namely that small amounts of blood lead do cause a few points of IQ loss. Then a question must be addressed: Is it reasonable to assume that the relationship between blood lead and IQ is a linear one that can be extrapolated down to a minute amount of lead in the blood, and extended upward to accommodate large amounts of blood lead? The answer, I believe, is "No."

To determine whether the IQ–lead level relationship is linear, the authors of the various studies would have needed to present adjusted IQs for children at each portion of the lead-level continuum. Most researchers have not done that. Those who have provided pertinent data have presented the data in scatter plots or bar graphs, and these pictorial representations do not seem to reflect linear relationships (even if the authors chose to interpret the relationships as linear).

Dietrich et al. (1993, Fig. 2) presented a line graph that shows the mean adjusted and unadjusted Performance IQ for four lead-level groups (0–10, >10–15, $\geq$15–20, and >20 μg/dl). Only the adjusted values are interpretable, and these show no meaningful difference among the first three groups (each averaging a Performance IQ of 90 ± 2). Only the most extreme lead group deviated from the other three (averaging about 85 on Performance IQ), suggesting a threshold effect (at about 20 μg/dl) rather than a linear relationship.

Bellinger et al. (1992, p. 858) offer a bar graph that presents adjusted WISC-R Full Scale IQs and K-TEA Battery Composite standard scores for the following groups: 0–4.9, 5.0–9.9, 10.0–14.9, and $\geq$15.0 μg/dl. The two groups with the lowest lead levels were indistinguishable from each other, averaging IQs of 118–120 and standard scores of about 119–122. Similarly, the two groups with the highest lead levels were indistinguishable from each other, each earning mean IQs of about 112 and mean standard scores of about 110. Again, a threshold effect (this time at about 10–15 μg/dl) is a more realistic explanation of the relationship than is a linear one.

Hatzakis et al. (1989, Fig. 5) included a line graph that presents unadjusted and adjusted WISC-R Full Scale IQs for the following groups: $\leq$14.9, 15.0–24.9, 25.0–34.9, 35.0–44.9, and $\geq$45.0 μg/dl. The lowest two groups did not differ meaningfully from each other, averaging adjusted IQs of 90 ± 1. Likewise, the highest three groups earned similar mean IQs of 85 ± 2. Once more, there was an apparent threshold, this time at about 25.0–34.9 μg/dl, considerably above the 10–20 μg/dl range that is usually deemed "low." Although the graph suggested linearity for the four groups with lead levels of 15 and above (adjusted mean IQs of about 91, 86, 84, and 83 with increasing lead level), there was decidedly no linearity for children with lead levels below 15 μg/dl.

Fulton et al. (1987, Fig. 1) presented a scatter plot for 10 lead-level groups that are defined by the log blood lead. The mean adjusted British Ability Scales (Elliott, Murray, & Pearson, 1979) score difference from the school mean is presented for each group. Although the authors draw a line of best fit through the points, visual inspection suggests no meaningful difference in the means for any of the samples; the values for nine samples (all but the lowest lead-level group) seem virtually identical to each other. These data suggest neither a threshold effect nor a linear relationship.

The graphs shown by the authors of the aforementioned studies indicate that if BLL truly affects IQ negatively, then there is likely a threshold effect to explain the relationship; there does not, however, appear to exist a documented linear relationship between lead level and adjusted IQ. Or if such a linear relationship exists, then it does so only at moderate levels of blood lead.

2.2. Interpreting fractions of an IQ point has no scientific meaning

Various federal documents include statements about 0.32 IQ point or 0.678 IQ point, but such statements have no meaning at all. Fractionating an IQ point is asking an IQ test to do something that it is just not equipped to do. It is a bit like stepping on a $30 bathroom scale and expecting it to give your weight to the nearest hundredth of a pound, or measuring your daughter's height with a 6-in. ruler and trying to record the result to the nearest tenth of an inch. The main difference between fractionating an IQ point and these other examples is that it is even less sensible to try to interpret a fraction of an IQ point than to anticipate incredible accuracy from a bathroom scale or ruler.

When Francis Galton (1869, 1883) first "invented" the IQ test he defined intelligence as an amalgam of sensory–motor functions such as reaction time, strength of pull, visual and auditory acuity, and so forth. He was able to measure these functions with amazing reliability and accuracy, sometimes to the nearest tenth or hundredth. The problem was that his so-called IQ test, though reliable, was subsequently shown to have no validity as a measure of intelligence. It took Alfred Binet et al. (Binet, 1903; Binet & Henri, 1895) to convince the scientific world that in order to measure something as complex as human intelligence, the measurements had to be complex, such as tests of reasoning, memory, and judgment; and whenever complex measures are used, the inevitable side effect is to have measurement error and other forms of construct-irrelevant variance. Binet's willingness to accept a certain amount of error in order to achieve validity may have been his most important contribution to science.

Even the best IQ tests, such as Wechsler's (1991, 1997) scales, have a band of error of about ± 3 points that surrounds each person's obtained IQ, and that band of error only affords about 68% accuracy. To be 90% or 95% certain that you have captured a person's "true" IQ within the band of error, that confidence interval must be expanded to ± 5 or ± 6 points.

These errors occur for a variety of reasons such as boredom, fatigue, luck (good or bad), rapport with the examiner, and so forth, and are a built-in aspect of every IQ assessment. Sometimes, errors occur in the scoring of tests due to carelessness, as noted previously (Kaufman, 1990, pp. 101–105; Kaufman, 1994, pp. 127–128), and sometimes scoring errors occur that cannot be helped. Three of the five WISC-R or WISC-III Verbal subtests that contribute to a child's Verbal IQ (Similarities, Comprehension, Vocabulary) have scoring systems that are subjective and require examiners to distinguish among scores of 0, 1, and 2 for each item. The scoring systems are illustrative, not exhaustive, and some responses are hard to classify, especially the fine distinctions between scores of 0 and 1 or between scores of 1 and 2. Similarly, many of the child's subjective responses to items on these three subtests need to be queried by the examiner if they are incomplete or ambiguous. Unfortunately, the guidelines for querying are incomplete and ambiguous, leading to considerable examiner differences in exactly which responses they question and how often, in general, they tend to query a child. Experienced scorers will differ from one another both in querying responses and in evaluating each response's merit; consequently, the Verbal IQs earned by a child will vary from examiner to examiner — sometimes by several points — just based on the administration and scoring decisions made on subjective items.

Furthermore, the Performance Scale has a similar kind of problem. Three of the five subtests that make up WISC-R or WISC-III Performance IQ (Picture Arrangement, Block Design, Object Assembly) allot up to 3 bonus points for quick, perfect performance on most items. However, the time a child takes to solve an item varies considerably from examiner to examiner. It is often unclear when a child has finished solving an item. Many children will not tell the examiner when they have finished, even though they are instructed to do so. It is a subjective decision when to stop the stopwatch. Some children may seem like they have stopped solving a puzzle, but they are actually thinking quietly and will suddenly rearrange all of the pieces. It is common to turn off the stopwatch and then have the child continue to complete a nonverbal item. Again, there is great variability in the speed of an examiner's "trigger" with the stopwatch, which affects the number of bonus points earned and can affect substantially the Performance IQ earned by the child.

Since the Full Scale IQ is composed of the Verbal IQ and Performance IQ in equal parts, it is evident that errors on either IQ will add error to the Full Scale IQ. These errors — or differences in administration and scoring decisions — are an unavoidable aspect of IQ test administration, even with highly trained and experienced examiners. As stated previously, random errors do not bias the outcome of the lead–IQ research studies; differences in examiner administration and scoring techniques, and any clerical errors they might make, do not compromise the interpretability of the data analyses. However, these prevalent differences among examiners, when interpreted in the context of the confidence interval of ± 5 or ± 6 points that surrounds everyone's "true" IQs, should underscore the lack of meaningfulness of a fraction of a single IQ point.

2.3. Societal consequences of lead level

Needleman (1989) and his associates (Needleman, Leviton, & Bellinger, 1982) have argued that as a result of elevated BLL in children, a much greater proportion of children will gravitate to the lower tail of the normal curve, earning low IQs, and fewer will earn high IQs. This inevitable ''shift'' will have a societal impact, burdening society with the need to provide increased special education services for the increased numbers of low-functioning children. The federal government has demonstrated concern about this societal shift, including statements in various guidelines (e.g., Federal Register, 1998) that stress ''avoided incidence of IQ below 70'' (p. 30321) and discuss how to compute economic benefits: ''The economic value of avoiding cases of IQ less than 70 is approximated by using avoided special education costs'' (p. 30321).

These societal implications derive, first, from the assumption that low BLLs really do produce IQ loss, that the relationship is linear, and it produces a constant effect across the age range (all of which may or may not be true); second, from the notion that IQ is the sole determinant of special education placement; third, with the implicit assumption that a person's IQ is somehow an unchangeable, absolute construct; and fourth, from the perspective that the loss of a few IQ points is surely related to diminished functioning within society. None of these assumptions or notions is true.

Certainly, IQs below 70 are not sufficient to place a child in a special education class. Mental retardation, by legal definition and diagnostic practice, cannot be diagnosed without evidence of retarded intellectual functioning and retarded adaptive functioning (Kamphaus, Reynolds, & Imperato-McCammon, 1999). Low IQs, by themselves, do not qualify a person for special education services; low adaptive behavior must accompany the cognitive retardation. I have seen no lead study that has examined a child's social-adaptive behavior with an instrument such as the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984), so we do not know the relationship between low BLL and adaptive behavior.

Additionally, there is no absolute definition of the kinds of cognitive behaviors that define IQs of 70 or 75 or 80. The IQ concept is not an absolute; it is not determined by a specific set of skills that indicate that a person is deficient in this or that type of mental functioning. Rather, low IQs are relative concepts. What defines an IQ of, say, 75 changes over time. As indicated in the discussion of generational changes in IQ, Americans are getting smarter at the rate of about 3 points per decade. That means that the same exact test performance that merited an IQ of 75 in 1960 (e.g., answering five questions of general information, solving three block designs, defining six vocabulary words) would only merit an IQ of 72 in 1970 and would now (near the year 2000) merit an IQ of only 63!

The yardstick for defining low, average, and high IQ is constantly changing. Test makers have to continue to determine what level of performance corresponds to different levels of IQ. Some variables may tend to lower a society's IQ (such as rate of unemployment), while others will counteract those variables and raise society's average IQ (such as the amount of information that can be accessed from a personal computer). Nevertheless, the end result is that the same percentage of people will wind up in each ''tail'' of the bell curve for the simple reason that IQ is an interval scale, not a ratio scale. Therefore, whenever test makers restandardize a test, they can identify the midpoint with accuracy, but the end points cannot actually be located. The only real choice is to have the obtained data for the new norms squeezed into the predictable shape of the bell curve. Those norms reflect the contemporary ''yardstick'' for equating test performance to IQ level, and the IQ distributions are made to fit the precise statistical properties of the normal curve; the percents in the tails do not change.

Imagine that scientists discovered an artificial neuronal transmitter that would increase everyone's IQ by 20 points. All of a sudden, the percentage of people in the "tail" at the high end of the curve would increase dramatically and there would be almost no one left in the tail at the low end. The mean would shift from 100 to 120. Nevertheless, all that would be temporary. The test developers would simply say, "Our norms are now very wrong. Let's go out and get a new standardization sample." They would do so, and then everything would be back to "normal." The tail at the high end would still produce about 2% who score above 130; except now, these are the people who would have scored 150 on the old norms; and the tail at the low end would still produce about 2% who earn IQs below 70. These people are now in the lower tail of the normal curve, and will perhaps need special placement. It makes no difference that before the artificial neuronal transmitter they would have earned IQs as high as 90. Now, they are retarded by the new, current definition of mentally retarded intellectual functioning. Categorization is thus relative to other people, not on an absolute scale.

IQs are relative concepts. Every time the norms get out of date, the publishers simply restandardize to get them back in line. It makes no difference whether variables are lowering or raising the IQs of its citizens. New bell curves are always being formed to reflect updated norms, and the percentages of individuals at different IQ levels can never really rise or fall; it will always be returned to the percents that define the normal curve. Needleman's (1989) societal arguments are illusory. IQ is a relative concept that is constantly in flux; there is nothing absolute about it.

The ultimate societal consequence is the loss of earnings that is indirectly attributed to the presence of low BLLs in young children. The line of reasoning is that low BLLs equal the loss of a few IQ points which, in turn, diminishes a child's ultimate earning capacity as an adult. As with the fragmentation of a single IQ point into its component parts, this societal equation is asking more than IQ tests were ever designed to deliver. From the beginning, they were developed as practical tools, not from any comprehensive theory of intelligence or cognition. Binet's (1903) quest, back in the late 1800s and early 1900s in Paris, was to identify school children who were likely to do poorly in school. His tests were largely verbal measures of memory, comprehension, judgment, and verbal expression. Nonverbal measures were added to the mix during World War I when new methods were needed to assess non-English speaking immigrants for service in the military. In the 1930s, David Wechsler (1939) blended the verbal approach of Binet with the nonverbal emphasis derived from the first World War, and the modern notion of intelligence — as measured by IQ tests — was born (Kaufman, 2000).

Contemporary IQ tests, therefore, trace their direct roots to Binet's work in France about a century ago and to the need to develop nonverbal tests to assess members of the Armed Forces in World War I. Binet's goal was to predict school achievement. The World War I psychologists were trying to evaluate the intelligence of people who did not speak English well, and, importantly, to detect malingerers. Those were the original goals for constructing the tasks that remain popular for assessing the IQ of children and adults. Because Wechsler's scales remain the most popular IQ tests today, tasks developed between a half-century and a century ago are the same subtests used today to make decisions about children and adults.

As a whole, the prediction of school-related ability is still a main goal of all current individually administered IQ tests; also, they have important clinical uses for identifying mentally retarded, learning disabled, and gifted children; adults with mental retardation, dyslexia, Alzheimer's disease; and adolescents and adults who need vocational or scholastic guidance (Kaufman, 1990, 1994; Reynolds & Gutkin, 1999). IQ tests are designed to measure cognitive problem-solving abilities and brain functioning, and are intended for use with a variety of people with known or suspected neurological problems, emotional or behavioral problems, school learning or memory problems, attention-deficit disorders, and the like.

IQ tests deliberately measure a limited aspect of human functioning. They are not intended to be used as the sole criterion for making any decisions that have educational, vocational, neurological, or societal implications. They are too narrow in scope and in design. They are not intended to measure interpersonal skills (social intelligence), creativity, special talents, or any of a number of qualities that are commonly associated with intelligent people. When they are used within schools, IQ tests perform reasonably well. Scores on IQ tests typically correlate about .50 to .70 with various criteria of school success. However, even coefficients of that magnitude explain about 25–50% of the variability in achievement scores, meaning that variables other than IQ account for one-half to three-quarters of the variability in school success; and once you get out of the school environment and into the workplace, the coefficients are even lower. IQ tests typically predict job success within the .20 to .40 range (Kaufman, 1990, Chapter 1). In addition, when you move out of the school and the workplace, correlations are even lower. As Anastasi and Urbina (1997) note: "For the prediction of practical criteria, not one but several tests may often be required. Most criteria are complex, the criterion measure depending on a number of different traits" (p. 156).

As indicated, neither Wechsler's tests nor the original Binet scale is theory-based, but evolved from practical considerations. Newer intelligence tests have evolved from theory, with two theories proving most influential: Luria's (1980) neuropsychological theory and Horn's (1989) and Horn and Cattell's (1966) fluid-crystallized theory of intelligence. In addition, the Luria and Horn frameworks have provided suitable theoretical foundations for better understanding children's and adult's patterns of scores on Wechsler's scales (Kaufman, 1990, 1994). Although the Luria and Horn theories are research based, they, nonetheless, offer a limited view of the world, focusing on one main area: How people solve problems. With only a few exceptions, subtests on Wechsler's scales and other contemporary intelligence tests more closely resemble laboratory tasks than the kinds of real-life problem solving that people are confronted with in their daily life. Or they resemble school-like tasks (answering general information, arithmetic, and vocabulary items), or they are game-like (putting together picture puzzles). However, the kinds of things individuals are asked to do on intelligence tests do not reflect everyday life. Subtests on IQ tests were generally constructed to be objective, easy to score, and straightforward. Real life is complex and intricate, and is not very conducive to such structure. It involves creative thought, social interactions, and much more than conventional IQ tests were ever intended to measure.

A more pertinent theory of intelligence for evaluating societal impact would be one that encompasses diverse aspects of intellectual functioning, such as Sternberg's (1985, 1997) triarchic theory of intelligence — a three-pronged theory that stresses analytic abilities, practical or adaptive skills, and insightful-thinking skills as components of human intelligence. He criticizes conventional IQ tests for measuring only one of the three essential prongs — analytical abilities. In contrast, Sternberg's (1993) unpublished group-administered test, the Sternberg Triarchic Abilities Test (STAT), attempts to measure all three components of his theory. Basically, Sternberg has broken down cognitive problem solving into a series of components that he has researched fairly extensively. His triarchic concept of intelligence involves applying these components to (a) abstract and academic problems (analytic thinking), (b) novel and unfamiliar problems (creative thinking), and (c) concrete and familiar everyday problems (practical thinking). IQ tests measure analytic thinking quite well, but not the other two types of intelligence. Creative thinking is required for many aspects of successful functioning in society, both in school, on the job, and in dealing with people. Practical thinking is associated in Sternberg's theory with "tacit knowledge," or the kinds of information and skills that people need to succeed in a variety of situations, for example, getting into the college of your choice, having a successful job interview, maintaining the respect of colleagues who work under you, and so forth. Taken together, the three types of intelligence, though not all-inclusive, would have many more societal implications than any one of the three in isolation. If an individual scored unusually low in all three areas due to a factor such as lead level, then that kind of finding might have implications for society in a more global sense. However, a person who might be handicapped in one arena could conceivably compensate in one or both of the other arenas.

In Sternberg's (1993, 1997) research, he has shown that the three components of intelligence correlate only modestly with each other, as each component has its unique aspects. A person who has low analytic ability (akin to a low IQ on a conventional IQ test) is just about as likely as a person with a high IQ to perform well on creative and practical tasks. In regression analyses intended to predict performance in high school courses, Sternberg (1993) showed that the creative and practical components consistently improved prediction significantly over and above the prediction that was obtained solely from the analytic component. In other words, even for academic courses, there is more to being successful than just having a high IQ. As currently measured, IQ is too narrow a concept to have societal implications even if exposure to very low levels of lead should be shown to lower IQ by a few IQ points. Theories of intelligence that are used to develop new IQ tests, or for interpreting Wechsler's tests, are much too limited in scope to affect a society as a whole. IQ tasks are not sufficiently real-world-oriented and do not tap an adequate breadth of mental abilities to encompass the kinds of activities that are necessary to maintain and advance a society. When other, more comprehensive, theories are applied (such as Sternberg's), and when future, broader-based intelligence tests are perfected (such as individually administered versions of the STAT), then perhaps the results of the lead–IQ research investigations will prove to have societal implications.

3. Summary

To summarize the essential shortcomings in the best-designed and executed lead IQ studies, we have the following.

(1) With such a small effect attributed to lead, uncontrolled variables cloud conclusions drawn from even the best studies. No matter how many confounding or contaminating

variables are controlled in the lead–IQ research studies, there are many additional variables — pertaining to subtle aspects of SES, childhood diseases, parenting skills, and even unknown influences on IQ — that remain uncontrolled in every lead–IQ study. Whenever lead level is found to be a significant predictor of IQ, after first controlling for several variables, researchers sometimes conclude that it is lead level, and lead level alone, that accounts for the significant loss of a few IQ points. A more correct statement is that any significant increase in prediction is due not only to lead level, but also to all other potentially important variables — known or unknown — that were not controlled in the study.

(2) Even in the best studies, parental IQ — a key variable affecting children's IQ — was either measured poorly or not at all. Parents' IQ relates to a variety of genetic and environmental factors that contribute substantially to their children's IQs and correlates significantly with their BLLs. This crucial variable has been recognized by many lead researchers, but the measurement of parents' IQ has sometimes been done poorly. Some researchers have not controlled for parents' IQ and others have used brief tests, such as tests of picture vocabulary, that are not truly measures of intelligence. Furthermore, mothers are invariably the only ones assessed in these studies; fathers are almost universally ignored.

(3) It is inappropriate for researchers to conduct many analyses at once, and then choose to interpret only the ones that support their position. Experimenters sometimes use what is known as a "shotgun" approach. They conduct a great number of analyses at once in an attempt to find at least one significant finding. Whenever this approach is used, chance error assumes a large role in the results (the more analyses that are conducted, the greater the likelihood that a significant result will emerge due to chance alone). The correct procedure is to apply a statistical correction to control for the error that occurs when many analyses are done at once. Several of the lead researchers have used this shotgun approach, but none have controlled for the known errors that accompany this approach.

(4) It is inappropriate to compare the IQs of the two extreme "lead-level" groups when several additional groups are included in the study. Several studies of lead and IQ determined the number of IQ points that can be attributed to slight elevations of lead level by comparing the IQs of the two extreme "lead-level" groups (the ones with the highest and lowest lead levels). In these studies, the researchers simply eliminated the middle groups and focused all attention on the extremes. That approach violates the rules for interpreting the results of multiple regression analysis, a procedure that is based on the total group of children; arbitrarily eliminating subgroups of children and focusing on the extremes, once again, takes advantage of chance errors. The number of IQ points loss attributed to lead, when based on "extreme group" procedures, is bogus.

(5) There is a lack of quality control in the administration and scoring of IQ tests. Individually administered tests of intelligence are clinical instruments that require careful, high-level training in order to yield valid results. In addition, even experienced examiners make clerical errors when scoring the tests. It is, therefore, important to use qualified examiners for the assessment of children's IQ and to incorporate quality control to ensure the validity of the data. Only a handful of lead–IQ researchers have recognized the subtleties involved in obtaining valid IQ data and have incorporated quality control procedures. In

contrast, several research teams seemed unaware of the need for trained examiners or of the need for quality control.

Despite the shortcomings in the best lead–IQ studies, the basic finding of an IQ loss, perhaps about 3 points in magnitude, has been accepted by many of the investigators as a well-validated finding. This finding has then been put to use in shaping public policy. However, there are some inappropriate practices that have been followed in the application of the research findings, as summarized here.

(1) Assumption of a linear relationship between lead level and IQ. There has been a tacit assumption that the relationship between BLL and IQ is a linear one that can be extrapolated down to a minute amount of lead, and extended upward to accommodate large amounts of lead. Yet, evaluation of the data that have been presented by lead researchers clearly does not support that assumption. When average IQs have been compared for groups that differ systematically in BLL, the best explanation of the relationship is as follows: If slight elevations of lead produce small decrements in IQ, then there seems to be a threshold effect; no cogent arguments can be made for a linear relationship.

(2) Interpretation of fractions of a single IQ point. The fragmentation of a single IQ point into fractions, as is commonly done in federal policy documents, is not meaningful. The IQs have a standard error of measurement of about ± 3 points, resulting from influences such as rapport with the examiner, fatigue, boredom, luck, and mood. In addition, different examiners make different administration decisions (e.g., when to query a response) and score subjective verbal items differently, leading to even greater error. The IQ tests are good, but they are not precise enough to permit an IQ point to be subdivided. Indeed, the bands of error surrounding the IQs earned by children of all ages is even larger than the small number of points attributed to slight elevations of lead levels.

(3) Societal impact of the IQ loss attributed to lead. Needleman et al. have claimed that the loss of a few IQ points will create a substantial increase in the numbers of individuals with low IQs, leading to great societal consequences. This contention is not true. IQ tests, by themselves, cannot be used to diagnose children as having mental retardation. Additionally, IQ is not absolute, determined by a specific set of skills that indicate that a person is deficient in this or that type of mental functioning. Instead, low IQs are relative concepts that define low or high functioning relative to how others of the same age perform on the same test items. The percents of children or adults who earn low IQs on tests of intelligence will remain a constant over time. Whether children are becoming smarter due to increased educational technology or less intelligent due to an impoverished environment or ingestion of lead, these changes from year to year will not change the proportion of individuals who score high or low on an IQ test. For a variety of reasons, the percentage of children earning IQs in a given range (e.g., below 70, above 125) stays the same every time an IQ test is restandardized. Furthermore, existing IQ tests were developed for practical, not theoretical purposes. The most pertinent theories that pertain to current IQ tests emphasize cognitive problem solving (analytic thinking), to the exclusion of real-life activities and other aspects of intelligence, such as the creative and practical components of Sternberg's triarchic theory. In order to have societal impact, lead level would need to be shown, conclusively, to negatively affect children's functioning in diverse dimensions of intellect.

Appendix

Appendix to Table 1

Reference	Type of study[a]	Pb measured in	Meta-analysis[b]
Baghurst, P. A., McMichael, A. J., Wigg, N. R., Vimpani, G. V., Robertson, E. F., Roberts, R. J., & Tong, S. (1992). Environmental exposure to lead and children's intelligence at the age of seven years: the Port Pirie cohort study. *New England Journal of Medicine, 327*, 1279–1284.	P	Blood	P/S, S
Dietrich, K. N., Berger, O. G., Succop, P. A., Hammond, P. B., & Bornschein, R. L. (1993). The developmental consequences of low to moderate prenatal and postnatal lead exposure: intellectual attainment in the Cincinnati lead study cohort following school entry. *Neurotoxicology and Teratology, 15*, 37–44.	P	Blood	P/S, S
Ernhart, C. B., Morrow-Tlucak, M., Wolf, A. W., Super, D., Drotar, D. (1989). Low level exposure in the prenatal and early preschool periods: intelligence prior to school entry. *Neurotoxicology and Teratology, 11*, 161–170.	P	Blood	P/S
Cooney, G., Bell, A., & Stavron, C. (1991). Low level exposures to lead and neurobehavioural development: the Sydney study at seven years. In *Heavy metals in the environment* (pp. 16–19). Edinburgh: CEP Consultants.	P	Blood	P/S
Bellinger, D. C., Stiles, K. M., & Needleman, H. L. (1992). Low-level lead exposure, intelligence and academic achievement: a long-term follow-up study. *Pediatrics, 90*, 855–861.	P	Blood	P/S, S
Hatzakis, A., Kokkevi, A., Katsouyanni, K., Maravelias, K., Salaminios, F., Kalandidi, A., Koutselinis, A., Stefanis, K., & Trichopouylos, D. (1987). Psychometric intelligence and attentional performance deficits in lead-exposed children. *Heavy Metals in the Environment, 1,* 204–206. For more detailed information, see Hatzakis A., Kokkevi A., Maravelias C., Katsouyanni K., Salaminios F., Kalandidi A., et al. (1989). Psychometric intelligence deficits in lead-exposed children. In M. A. Smith, L. D. Grant, & A. I. Sora (Eds.) *Lead exposure and child development* (pp. 211–223). London: Kluwer.	C	Blood	P/S, N, S

Fulton, M., Raab, G., Thomson, G., Laxen, D., Hunter, R., & Hepburn, W. (1987). Influence of blood lead on the ability and attainment of children in Edinburgh. *Lancet, 1* (Saturday, 30 May), 1221–1226.	C	Blood	P/S, N, S
Winneke, G., Brockhaus, A., Ewers, U., Kramer, U., & Neuf, M. (1990). Results from the European multicenter study on lead neurotoxicity in children: implications for risk assessment. *Neurotoxicology and Teratology, 12*, 553–559.	C	Blood	P/S
Silva, P. A., Hughes, P., Williams, S., & Faed, J. M. (1988). Blood lead, intelligence, reading attainment, and behaviour in eleven year old children in Dunedin, New Zealand. *Journal of Child Psychology and Psychiatry, 29*, 43–52.	C	Blood	P/S, S
Yule, W., Lansdown, R., Millar, I. B., & Urbanowicz, M. A., (1981). The relationship between blood lead concentrations, intelligence and attainment in a school population: a pilot study. *Devevelopmental Medicine and Child Neurology, 23*, 567–576.	C	Blood	P/S, N, S
Lansdown, R., Yule, W., Urbanowicz, M. A., & Hunter, J. (1986). The relationship between blood-lead concentrations, intelligence, attainment and behaviour in a school population: the second London study. *International Archives Occupational and Environmental Health, 57*, 225–235.	C	Blood	P/S, N
Harvey, P. G., Hamlin, M. W., Kumar, R., Morgan, J., Spurgeon, A., & Delves, H. T. (1988). Relationships between blood lead, behaviour, psychometric and neuropsychological test performance in young children. *British Journal of Developmental Psychology, 6*, 145–156.	C	Blood	P/S
Wang, L., Xu, S., Thang, G., & Wang, W. (1989). Study of lead absorption and its effect on children's development. *Biomedical and Environmental Science, 2*, 325–330.	C	Blood	P/S
Ernhart, C. B., Landa, B., & Wolf, A. W. (1985). Subclinical lead level and developmental deficit: reanalysis of data. *Journal of Learning Disabilities, 18*, 475–479.	C	Blood	N

(continued on next page)

Appendix (continued)

Reference	Type of study[a]	Pb measured in	Meta-analysis[b]
Schroeder, S. R., Hawk, B., Otto, D. A., Mushak, P., & Hicks, R. E. (1985). Separating the effects of lead and social factors on IQ. *Environmental Research, 38*, 144–154.	C	Blood	N
Hawk, B. A., Schroeder, S. R., Robinson, G., Otto, D., Mushak, P., Kleinbaum, D., & Dawson, G. (1986). Relation of lead and social factors to IQ of low-SES children: a partial replication. *American Journal of Mental Deficiency, 91*, 178–183.	C	Blood	N, S
Winneke, G., Beginn, U., Ewert, T., Havestadt, C., Kraemer, U., Krause, C., Thron, H. L., & Wagner, H. M. (1985). Comparing the effects of perinatal and later childhood lead exposure on neuropsychological outcome. *Environmental Research, 38*, 155–167.	C	Blood	P/S
Fergusson, D. M., Fergusson, J. E., Horwood, L. J., & Kinzett, N. G. (1988). A longitudinal study of dentine lead levels, intelligence, school performance and behaviour: Part II. Dentine lead and cognitive ability. *Journal of Child Psychology and Psychiatry, 29*, 793–809.	C	Tooth	P/S, N
Smith, M., Delves, T., Lansdown, R., Clayton, B., & Graham, P. (1983). The effects of lead exposure on urban children: the Institute of Child Health/ Southhampton study. *Developmental Medicine and Child Neurology, 25* (Suppl. 47) 1–54.	C	Tooth	P/S
McMichael, A. J., Baghurst, P. A., Vimpani, G. V., Wigg, N. R., Robertson, E. F., & Tong, S. (1994). Tooth lead levels and IQ in school-age children: the Port Pirie cohort study. *American Journal of Epidemiology, 140*, 489–499.	C	Tooth	P/S
Fulton, M., Paterson, L., Raab, G., Thomson, G., & Laxen, D. (1989). Blood lead, tooth lead and child development in Edinburgh. In J. P. Vernet (Ed.), *Heavy metals in the environment* (vol. 2) (pp. 68–71). Edinburgh: CEP Consultants.	C	Tooth	P/S

Needleman, H. L., Gunnoe, C., Leviton, A., Reed, R., Pesesie, H., Maher, C., & Barrett, P. (1979). Deficits in psychologic and classroom performance of children with elevated dentine lead levels. *New England Journal of Medicine, 300*, 689–695.	C	Tooth	P/S
Winneke, G., Kramer, U., Brockhaus, A., Ewers, U., Kujanek, G., Lechner, H., & Janke, W. (1983). Neuropsychological studies in children with elevated tooth-lead concentrations: II. Extended study. *International Archives of Occupational and Environmental Health, 51*, 231–252.	C	Tooth	P/S, N
Bergomi, M, Borella, P., Fantuzzi, G., Vivoli, G., Sturloni, N., Cavazzuti, G., Tampieri, A., & Tartoni, P. L. (1989). Relationship between lead exposure indicators and neuropsychological performance in children. *Developmental Medicine and Child Neurology, 31*, 181–190.	C	Tooth	P/S
Pocock, S. J., Ashby, D., & Smith, M. A. (1987). Lead exposure and children's intellectual performance. *International Journal of Epidemiology, 16*, 57–67.	C	Tooth	N
Hansen, O. N., Trillingsgaard, A., Beese, I., Lyngbye, T., & Grandjean, P. (1989). A neuropsychological study of children with elevated dentine lead level: assessment of the effect of lead in different socio-economic groups. *Neurotoxicology and Teratology, 11*, 205–213.	C	Tooth	N

[a] P = Prospective; C = Cross-sectional.

[b] Study included in one or more of the following meta analyses: P/S = Pocock, Smith; Pocock, S. J. , Smith M., & Baghurst P. (1994). Environmental lead and children's intelligence: a systematic review of epidemiological evidence. *British Medical Journal 309*, 1189–1197. N = Needleman; Needleman, H. L., & Gatsonis, C. A. (1990). Low-level exposure and the IQ of children. *JAMA, the Journal of the American Medical Association 263*, 673–678. S = Schwartz; Schwartz, J. (1994). Low-level lead exposure and children's IQ: a meta analysis and search for a threshold. *Environmental Research 65*, 45–55.

References

Abidin, R. (1986). *Parenting Stress Index—manual (PSI)* (2nd ed.). Charlottesville, VA: Pediatric Psychology Press.

American Psychological Association. (1985). *Standards for educational and psychological testing*. Washington, DC: Author.

Ammons, R. B., & Ammons, C. H. (1962). The quick test (QT): provisional manual. *Psychological Reports*, *11*, 111–161.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Atkinson, S. E., Crocker, T. D., & Needleman, H. L. (1987). The importance of specification uncertainty and intolerance to measurement error in a study of the impact of dentine lead on children's IQs. *International Journal of Environmental Studies*, *29*, 127–138.

Baghurst, P. A., McMichael, A. J., Wigg, N. R., Vimpani, G. V., Robertson, E. F., Roberts, R. J., & Tong, S.-L. (1992). Environmental exposure to lead and children's intelligence at the age of seven years: the Port Pirie cohort study. *New England Journal of Medicine*, *327*, 1279–1284.

Baldwin, R. L. (1993). Effects of otitis media on child development. *The American Journal of Otology*, *14*, 601–604.

Banks, E. C., Ferretti, L. E., & Shucard, D. W. (1997). Effects of low level exposure on cognitive function in children: a review of behavioral, neuropsychological and biological evidence. *Neurotoxicology*, *18*, 237–282.

Bellinger, D., Leviton, A., Needleman, H. L., Waternaux, C., & Rabinowitz, M. (1986). Low-level lead exposure and infant development in the first year. *Neurobehavioral Toxicology and Teratology*, *8*, 151–161.

Bellinger, D., Leviton, A., & Waternaux, C. (1989). Lead, IQ and social class. *International Journal of Epidemiology*, *18*, 180–185.

Bellinger, D., Leviton, A., Waternaux, C., Needleman, H. L., & Rabinowitz, M. (1987). Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. *New England Journal of Medicine*, *316*, 1037–1043.

Bellinger, D. C., & Needleman, H. L. (1983). Lead and the relationship between maternal and child intelligence. *Pediatrics*, *102*, 523–527.

Bellinger, D., Needleman, H. L., Bromfield, R., & Mintz, M. (1984). A follow-up study of the academic attainment and classroom behavior of children with elevated dentine lead levels. *Biological Trace Element Research*, *6*, 207–223.

Bellinger, D., Sloman, J., Leviton, A., Rabinowitz, M., Needleman, H. L., & Waternaux, C. (1991). Low-level lead exposure and children's cognitive function in the preschool years. *Pediatrics*, *87*, 219–227.

Bellinger, D. C., Stiles, K. M., & Needleman, H. L. (1992). Low-level lead exposure, intelligence and academic achievement: a long-term follow-up study. *Pediatrics*, *90*, 855–861.

Bergomi, M., Borella, P., Fantuzzi, G., Vivoli, G., Sturloni, N., Cavazzuti, G., Tampieri, A., & Tartoni, P. L. (1989). Relationship between lead exposure indicators and neuropsychological performance in children. *Developmental Medicine and Child Neurology*, *31*, 181–190.

Binet, A. (1903). *L'etude experimentale de l'intelligence*. Paris: Schleicher.

Binet, A., & Henri, V. (1895). La psychologie individuelle. *L'Annee Psychologique*, *2*, 411–465.

Black, M. M., & Sonnenschein, S. (1993). Early exposure to otitis media: a preliminary investigation of behavioral outcome. *Journal of Developmental and Behavioral Pediatrics*, *14*, 150–155.

Caldwell, B., & Bradley, R. (1984). *Administration manual: home observation for measurement of the environment* (Rev. ed.). Little Rock, AR: University of Arkansas at Little Rock.

Centers for Disease Control. (1991). Preventing lead poisoning in young children. In: *A statement by the Centers for Disease Control* p. 3. Environmental Health Services Division, Atlanta, GA: Bureau of State Services.

Chandler, L., The Children's Life-Events Inventory: Instructions for Use. Unpublished manual, University of Pittsburgh, Pittsburgh, PA, 1982.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*, 98–101.

Cooney, G., Bell, A., & Stavron, C. (1991). Low level exposures to lead and neurobehavioural development: the Sydney study at seven years. In: *Heavy metals in the environment* (pp. 16–19). Edinburgh, UK: CEP Consultants.

Dietrich, K. N., Berger, O. G., Succop, P. A., Hammond, P. B., & Bornschein, R. L. (1993). The developmental consequences of low to moderate prenatal and postnatal lead exposure: intellectual attainment in the Cincinnati lead study cohort following school entry. *Neurotoxicology and Teratology*, *15*, 37–44.

Dunn, L. (1965). *Manual for the Peabody Picture Vocabulary Test (PPVT)*. Circle Pines, MN: American Guidance Service.

Dunn, L., & Dunn, L. (1981). *Manual for the Peabody Picture Vocabulary Test-Revised (PPVT-R)*. Circle Pines, MN: American Guidance Service.

Elliott, C. D., Murray, D. J., & Pearson, L. S. (1979). *British ability scales*. Windsor, UK: NFER/Nelson.

Ernhart, C. B. (1992). A critical review of low level prenatal lead exposure in the human: 1. Effects on the fetus and newborn. *Reproductive Toxicology*, *6*, 9–19.

Ernhart, C. B. (1995). Inconsistencies in the lead-effects literature exist and cannot be explained by "effect modification". *Neurotoxicology and Teratology*, *17*, 227–233.

Ernhart, C. B., Landa, B., & Wolf, A. W. (1985). Subclinical lead level and developmental deficit: reanalysis of data. *Journal of Learning Disabilities*, *18*, 475–479.

Ernhart, C. B., Morrow-Tlucak, M., Wolf, A. W., Super, D., & Drotar, D. (1989). Low level exposure in the prenatal and early preschool periods: intelligence prior to school entry. *Neurotoxicology and Teratology*, *11*, 161–170.

Federal Register (1996). Vol. 61, No. 111, Friday, June 7, 1996.

Federal Register (1998). Vol. 63, No. 106, Wednesday, June 3, 1998.

Fergusson, D. M., Fergusson, J. E., Horwood, L. J., & Kinzett, N. G. (1988). A longitudinal study of dentine lead levels, intelligence, school performance and behaviour: Part II. Dentine lead and cognitive ability. *Journal of Child Psychology and Psychiatry*, *29*, 793–809.

Fischler, R. S., Todd, W., & Feldman, C. M. (1985). Otitis media and language performance in a cohort of Apache Indian children. *American Journal of Diseases of Children*, *139*, 355–360.

Flaherty, J., Gaviria, F., & Pathak, D. (1983). The measurement of social support: the social support network inventory. *Comprehensive Psychiatry*, *24*, 521–529.

Flynn, J. R. (1983). Now the great augmentation of the American IQ. *Nature*, *301*, 655.

Flynn, J. R. (1984). The mean IQ of Americans: massive gains 1932–1978. *Psychological Bulletin*, *95*, 29–51.

Flynn, J. R. (1987). Massive gains in 14 nations: what IQ tests really measure. *Psychological Bulletin*, *101*, 171–191.

Friel-Patti, S., & Finitzo, T. (1990). Language learning in a prospective study of otitis media with effusion in the first two years of life. *Journal of Speech and Hearing Research*, *33*, 188–194.

Fulton, M., Paterson, L., Raab, G., Thomson, G., & Laxen, D. (1989). Blood lead, tooth lead and child development in Edinburgh. In: J. P. Vernet (Ed.), *Heavy metals in the environment*, *vol. 2* (pp. 68–71). Edinburgh U.K: CEP Consultants.

Fulton, M., Raab, G., Thomson, G., Laxen, D., Hunter, R., & Hepburn, W. (1987). Influence of blood lead on the ability and attainment of children in Edinburgh. *Lancet*, *1*, 1221–1226 (Saturday, 30 May).

Galton, F. (1869). *Hereditary genius: an inquiry into its laws and consequences*. London: Macmillan.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Goldstein, G. (1984). Brain capillaries: a target for inorganic lead poisoning. *Neurotoxicology*, *5*, 167–176.

Greene, T., & Ernhart, C. B. (1993). Dentine lead and intelligence prior to school entry: a statistical sensitivity analysis. Response. *Journal of Clinical Epidemiology*, *46*, 404–406.

Gregory, R. J. (1987). *Adult intellectual assessment*. Boston, MA: Allyn and Bacon.

Gregory, R. J., Lehman, R. E., & Mohan, P. J. (1976). Intelligence test results for children with and without undue lead absorption. In: G. Wegner (Ed.), *Shoshone lead health project* (pp. 120–150). Boise, ID: Idaho Department of Health and Welfare.

Hansen, O. N., Trillingsgaard, A., Beese, I., Lyngbye, T., & Grandjean, P. (1989). A neuropsychological study of children with elevated dentine lead level: assessment of the effect of lead in different socio-economic groups. *Neurotoxicology and Teratology*, *11*, 205–213.

Harsten, G., Nettelbladt, U., Schalen, L., Kalm, O., & Prellner, K. (1993). Language development in children with acute otitis media during the first three years of life. Follow-up study from birth to seven years. *Journal of Laryngology and Otology*, *107*, 407–412.

Harvey, P. G., Hamlin, M. W., Kumar, R., Morgan, J., Spurgeon, A., & Delves, H. T. (1988). Relationships between blood lead, behaviour, psychometric and neuropsychological test performance in young children. *British Journal of Developmental Psychology*, *6*, 145–156.

Hatzakis, A., Kokkevi, A., Katsouyanni, K., Maravelias, K., Salaminios, F., Kalandidi, A., Koutselinis, A., Stefanis, K., & Trichopouylos, D. (1987). Psychometric intelligence and attentional performance deficits in lead-exposed children. *Heavy Metals in the Environment*, *1*, 204–209.

Hatzakis, A., Kokkevi, A., Maravelias, C., Katsouyanni, K., Salaminios, F., Kalandidi, A., Koutselinis, A., Stefanis, C., & Trichopolous, D. (1989). Psychometric intelligence deficits in lead-exposed children. In: M. A. Smith, L. D. Grant, & A. L. Sora (Eds.), *Lead exposure and child development* (pp. 211–223). London: Kluwer.

Hawk, B. A., Schroeder, S. R., Robinson, G., Otto, D., Mushak, P., Kleinbaum, D., & Dawson, G. (1986). Relation of lead and social factors to IQ of low-SES children: a partial replication. *American Journal of Mental Deficiency*, *91*, 178–183.

Holm, V. A., & Kunze, L. H. (1969). Effect of otitis media on language development. *Pediatrics*, *43*, 833–839.

Holmes, T., & Raye, R. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, *11*, 213–218.

Horn, J. L. (1989). Cognitive diversity: a framework of learning. In: P. L. Ackerman, J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 61–116). New York: Freeman.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253–270.

Kamphaus, R. W., Reynolds, C. R., & Imperato-McCammon, C. (1999). Roles of diagnosis and classification in school psychology. In: C. R. Reynolds, & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed.) (pp. 292–306). New York: Wiley.

Katz, J. (1978). The effects of conductive hearing loss on auditory function. *ASHA*, *10*, 879–886.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn and Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.

Kaufman, A. S. (2000). Tests of intelligence. In: R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 445–476). New York: Cambridge Univ. Press.

Kaufman, A. S., & Kaufman, N. L. (1983). *Manual for the Kaufman Assessment Battery for Children (K-ABC)*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1985). *Manual for the Kaufman Test of Educational Achievement (K-TEA) brief form*. Circle Pines, MN: American Guidance Service.

Kotok, D. (1972). Development of children with elevated blood levels: a controlled study. *Journal of Pediatrics*, *80*, 57–61.

Landrigan, P., Balow, R., Whitworth, R., Staeling, N., & Rosenbloom, B. (1975). Neuropsychological dysfunction in children with chronic low level lead absorption. *Lancet*, *1*, 708.

Lansdown, R., Yule, W., Urbanowicz, M. A., & Hunter, J. (1986). The relationship between blood-lead concentrations, intelligence, attainment and behaviour in a school population: the second London study. *International Archives Occupational Environmental Health*, *57*, 225–235.

Lehman, R. S. (1991). *Statistics and research design in the behavioral sciences*. Belmont, CA: Wadsworth.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York: Basic Books.

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, *297*, 222–223.

Lynn, R. (1990). The role of nutrition in secular increases in intelligence. *Personality and Individual Differences*, *11*, 273–285.

McLean, J. E. (1995). *Improving education through action research*. Thousand Oaks, CA: Corwin Press.

McMichael, A. J., Baghurst, P. A., Vimpani, G. V., Wigg, N. R., Robertson, E. F., & Tong, S. (1994). Tooth lead levels and IQ in school-age children: the Port Pirie cohort study. *American Journal of Epidemiology*, *140*, 489–499.

McShane, D., & Plas, J. (1984). The cognitive functioning of American Indian children. *School Psychology Review*, *13*, 61–73.

Needleman, H. L. (1982). Lead-associated intellectual deficit. *New England Journal of Medicine*, *306*, 367.

Needleman, H. L. (1987). Low level lead exposure in the fetus and young child. *Neurotoxicology*, *8*, 389–394.

Needleman, H. L. (1989). The persistent threat of lead: a singular opportunity. *American Journal of Public Health*, *79*, 643–645.

Needleman, H. L., & Gatsonis, C. A. (1990). Low-level lead exposure and the IQ of children, meta-analysis of modern studies. *Journal American Medical Association*, *263*, 673–678.

Needleman, H. L., Geiger, S. K., & Frank, R. (1983). Lead and IQ scores: a reanalysis. *Science*, *227*, 701–703.

Needleman, H. L., Gunnoe, C., Leviton, A., Reed, R., Pesesie, H., Maher, C., & Barrett, P. (1979). Deficits in psychologic and classroom performance of children with elevated dentine lead levels. *New England Journal of Medicine*, *300*, 689–695.

Needleman, H. L., Leviton, A., & Bellinger, D. (1982). Lead-associated intellectual deficit. *New England Journal of Medicine*, *306*, 367.

Needleman, H. L., Riess, J. A., Tobin, M. J., Biesecker, G. E., & Greenhouse, J. B. (1996). Bone lead levels and delinquent behavior. *Journal American Medical Association*, *275*, 363–369.

Olson, D., Portner, L., Lavee, Y. (1985). Family Adaptability and Cohesive Evaluation Scales-III. Unpublished manual, University of Minnesota, St. Paul, MN.

Pocock, S. J., Ashby, D., & Smith, M. (1987). Lead exposure and children's intellectual performance. *International Journal of Epidemiology*, *16*, 57–67.

Pocock, S. J., Smith, M., & Baghurst, P. (1994). Environmental lead and children's intelligence: a systematic review of the epidemiological evidence. *British Medical Journal*, *309*, 1189–1197.

Rach, G. H., Zeilhuis, G. A., & Van den Broek, P. (1988). The influence of chronic persistent otitis media with effusion on language development of 2 to 4 year olds. *International Journal of Pediatric Otorhinology*, *15*, 253–261.

Residential Lead-Based Paint Hazard Reduction Act of 1992 (1992). Title X of the Housing and Community Development Act of 1992, October 12, 1992. Pub. L 102–550. (Provisions of Title X are codified in U.S.C., vol. 42, sec. 4851.)

Reynolds, C. R. (1999). The inference of causality between smoking and low birth weight: good science or good politics? *Journal of Forensic Neuropsychology, 1*, 55–87.

Reynolds, C. R., & Gutkin, T. B. (Eds.) (1999) *The handbook of school psychology* (3rd ed., pp. 292–306). New York: Wiley.

Roberts, J. E., Burchinal, M. R., Davis, B. P., Collier, A. M., & Henderson, F. W. (1991). Otitis media in early childhood and later language. *Journal of Speech and Hearing Research*, *34*, 1158–1168.

Robertson, G. J., & Eisenberg, J. L. (1981). *PPVT-R technical supplement*. Circle Pines MN: American Guidance Service.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist*, *45*, 775–778.

Rosenthal, R. (1991). Effect sizes: Pearson's correlation, its display via the BESD, and alternative indices. *American Psychologist*, *46*, 1086–1087.

Sachs, H. K. (1996). Bone lead levels and delinquent behavior — to the editor. *Journal American Medical Association*, *275*, 1725–1726.

Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego: Sattler.

Sayre, J. W. (1996). Bone lead levels and delinquent behavior — to the editor. *Journal American Medical Association*, *275*, 1726.

Schroeder, S. R., Hawk, B., Otto, D. A., Mushak, P., & Hicks, R. E. (1985). Separating the effects of lead and social factors on IQ. *Environmental Research*, *38*, 144–154.

Schwartz, J. (1994). Low level lead exposure and children's IQ: a meta-analysis and search for a threshold. *Environmental Research*, *65*, 42–55.

Silva, P. A., Hughes, P., Williams, S., & Faed, J. M. (1988). Blood lead, intelligence, reading attainment, and behaviour in eleven year old children in Dunedin, New Zealand. *Journal of Child Psychology and Psychiatry*, *29*, 43–52.

Smith, M., Delves, T., Lansdown, R., Clayton, B., & Graham, P. (1983). The effects of lead exposure on urban children: the Institute of Child Health/Southampton study. *Developmental Medicine and Child Neurology*, *25* (Suppl. 47), 1–54.

Sonderegger, T. B. (Ed.). (1992) *Perinatal substance abuse: research findings and clinical implications*. Baltimore, MD: Johns Hopkins Univ. Press.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales, expanded form manual*. Circle Pines, MN: American Guidance Service.

Sternberg, R. J. (1985). *Beyond IQ: a triarchic theory of human intelligence*. New York, NY: Cambridge Univ. Press.

Sternberg, R.J. (1993). Sternberg Triarchic Abilities Test (STAT). Unpublished test.

Sternberg, R. J. (1997). Educating intelligence: infusing the triarchic theory into school instruction. In: R. J. Sternberg, & E. Sternberg (Eds.), *Intelligence, heredity, and environment* (pp. 343–362). New York, NY: Cambridge Univ. Press.

Teele, D. W., Klein, J. O., Chase, C., Menyuk, P., & Rosner, B. A. (1990). Otitis media in infancy and intellectual ability, school achievement, speech, and language at age 7 years. *The Journal of Infectious Diseases*, *162*, 685–694.

Teele, D. W., Klein, J. O., & Rosner, B. A. (1984). Otitis media with effusion during the first three years of life and development of speech and language. *Pediatrics*, *74*, 282–287.

Tong, S., Baghurst, P., McMichael, A., Sawyer, M., & Mudge, J. (1996). Lifetime exposure to environmental lead and children's intelligence at 11–13 years: the Port Pirie cohort study. *British Medical Journal*, *312*, 1569–1575.

Updike, C., & Thornburg, J. D. (1992). Reading skills and auditory processing ability in children with chronic otitis media in early childhood. *Annals of Otology, Rhinology, and Laryngology*, *101*, 530–537.

Van Baar, A. (1990). Development of infants of drug dependent mothers. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *31*, 911–920.

Wang, L., Xu, S., Thang, G., & Wang, W. (1989). Study of lead absorption and its effect on children's development. *Biomedical and Environmental Sciences*, *2*, 325–330.

Wasserman, G. A., Graziano, J. H., Factor-Litvak, P., Popovac, D., Morina, N., Musabegovic, A., Vrenezi, N., Capuni-Paracka, S., Lekic, V., Preteni-Redjepi, E., Hadzialjevic, S., Slavkovich, V., Kline, J., Shrout, P., & Stein, Z. (1994). Consequences of lead exposure and iron supplementation on childhood development at age 4 years. *Neurotoxicology and Teratology*, *16*, 233–240.

Wasserman, G. A., Liu, X., Lolacono, N. J., Factor-Litvak, P., Kline, J., Popovac, D., Morina, N., Musabegovic, A., Vrenezi, N., Capuni-Paracka, S., Lekic, V., Preteni-Redjepi, E., Hadzialjevic, S., Slavkovich, V., & Graziano, J. H. (1997). Lead exposure and intelligence in 7-year-old children. *Environmental Health Perspectives*, *105*, 956–962.

Wasserman, G. A., Staghezza-Jaramillo, B., Shrout, P., Popovac, D., & Graziano, J. (1998). The effect of lead exposure on behavior problems in preschool children. *American Journal of Public Health*, *88*, 481–486.

Webster, A., Bamford, J. M., Thyer, N. J., & Ayles, R. (1989). The psychological, educational and auditory sequelae of early, persistent secretory otitis media. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *30*, 529–546.

Wechsler, D. (1939). *Measurement of adult intelligence*. Baltimore, MD: Williams and Wilkins.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale (WAIS)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised (WISC-R)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised (WAIS-R)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III)*. San Antonio, TX: The Psychological Corporation.

WHO (1984). Consultation on monitoring of lead neurotoxicity in children. Executive Summary, European Regional Programme on Chemical Safety, Dusseldorf, Germany (June).

Winneke, G., Beginn, U., Ewert, T., Havestadt, C., Kraemer, U., Krause, C., Thron, H. L., & Wagner, H. M. (1985). Comparing the effects of perinatal and later childhood lead exposure on neuropsychological outcome. *Environmental Research*, *38*, 155–167.

Winneke, G., Brockhaus, A., Ewers, U., Kramer, U., & Neuf, M. (1990). Results from the European multicenter study on lead neurotoxicity in children: implications for risk assessment. *Neurotoxicology and Teratology*, *12*, 553–559.

Winneke, G., Kramer, U., Brockhaus, A., Ewers, U., Kujanek, G., Lechner, H., & Janke, W. (1983). Neuropsychological studies in children with elevated tooth-lead concentrations: II. Extended study. *International Archives Occupational Environmental Health*, *51*, 231–252.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson revised, tests of cognitive ability: standard and supplemental batteries*. Chicago: Riverside.

Yule, W., Lansdown, R., Millar, I. B., & Urbanowicz, M. A. (1981). The relationship between blood lead concentrations, intelligence and attainment in a school population: a pilot study. *Developmental Medicine and Child Neurology*, *23*, 567–576.

Yule, W., Urbanowicz, M. A., Lansdown, R., & Millar, I. B. (1984). Teachers' ratings of children's behaviour in relation to blood lead levels. *British Journal of Developmental Psychology*, *2*, 295–305.