A Multilevel Approach to the Relationship Between Birth Order and Intelligence

Aaron L. Wichman
The Ohio State University
Joseph Lee Rodgers
University of Oklahoma
Robert C. MacCallum
University of North Carolina–Chapel Hill

Many studies show relationships between birth order and intelligence but use cross-sectional designs or manifest other threats to internal validity. Multilevel analyses with a control variable show that when these threats are removed, two major results emerge: (a) birth order has no significant influence on children's intelligence and (b) earlier reported birth order effects on intelligence are attributable to factors that vary between, not within, families. Analyses on 7- to 8- and 13- to 14-year-old children from the National Longitudinal Survey of Youth support these conclusions. When hierarchical data structures, age variance of children, and within-family versus between-family variance sources are taken into account, previous research is seen in a new light.

Keywords: *hierarchical; nested; intelligence; multilevel; birth order; NLSY*

The topic of birth order may have drawn more attention from a wider variety of psychologists for a longer time than any other topic in our field. Psychologists interested in social processes, in development, in the family, in children, in reproduction, and in methodology have participated in birth order research. Other disciplines also have investigated birth order in a serious way, including anthropologists, sociologists, demographers, medical researchers, and even financial historians. Sir Francis Galton (1874) may have published the first research on birth order and intellectual ability (he noted a disproportionate number of first-borns among British scientists) but the topic of birth order has ancient status. For example, the birth order of Cain and Abel was relevant to the original Bible story. Even if Adam and Eve lacked training in psychology or research methods, they must have been among the very early observers of the effects of birth order patterns.

However, birth order is not easy to study. The methodological difficulties of properly accounting for effects of birth order are belied by its apparent simplicity. As Schooler (1972) noted, "it may well have been the seeming simplicity of birth order as an independent variable that provides the answer to . . . its attractiveness to researchers" (p. 174). Methodological critiques have been published at regular intervals, including Kammeyer (1967), Adams (1972), Schooler (1972), Schvaneveldt and Ihinger (1979), Ernst and Angst (1983), Rodgers and Thompson (1985), and Sulloway (1996).

Birth order patterns have been observed in relation to a wide array of dependent variables. Rodgers and Thompson (1985) noted that "birth order . . . is expected to predict the behavior of almost anyone: strippers and presidents, dentists and soldiers, assassins, authors, athletes, alcoholics, adult smokers, and assorted others" (p. 158). These topics notwithstanding, the most attention by far in the research literature has been given to the relation between birth order and intelligence.

Authors' Note: A special thanks goes to Dr. Stephen DuToit for his invaluable assistance. We also thank the staff of the Center for Human Resource Research of the Ohio State University for their help identifying variables for analysis. Correspondence concerning this article may be addressed to the first or second author. Contact Aaron L. Wichman at The Ohio State University, Department of Psychology, 122 Townshend Hall, Columbus, OH 43210-1222; e-mail: wichman.3@osu.edu. Contact Joseph Lee Rodgers at the Department of Psychology, University of Oklahoma, Norman, OK 73019; e-mail: jrodgers@ou.edu.

PSPB, Vol. 32 No. 1, January 2006 117-127
DOI: 10.1177/0146167205279581

Early attention to birth order and intelligence was impressive from both theoretical and methodological perspectives. Studies by Galton (1874), Thurstone and Jenkins (1929), Outhit (1933), Koch (1954), and Anastasi (1956) are exemplary of the early attention that was paid to the relation between birth order and intelligence by prominent psychologists. Unfortunately, even with such talent, coherence was lacking in the research: Kammeyer (1967) described the state of the birth order research enterprise at that time as "best characterized as a disparate, disconnected, aggregation of research findings" (p. 73).

In 1973, Belmont and Marolla published a study of the relation between birth order and intelligence from which the birth order literature has not yet recovered. The Belmont and Marolla study was a simple empirical compilation of Raven Progressive Matrices scores from a cross-section of almost 400,000 Dutch men of different birth orders. When the IQ scores were disaggregated by levels of birth order and family size, a remarkably systematic pattern emerged, which suggested declining intelligence with increasing birth order and family size. The belief that large families and later birth order had a causal influence on intelligence preceded the Belmont and Marolla study. Nevertheless, the graphical patterns in their study were so clean, and the sample size so large, that many viewed their results as the "final accounting." Belmont and Marolla cautioned that the differences, although highly systematic, "appear to be small." But the caution was not heeded. For example, in a published interview with Robert Zajonc, Elizabeth Hall (1986) noted that "the sheer volume of data (400,000 observations) convinced Zajonc . . . that the findings were no accident" (p. 47).

Since then, both the scientific and popular presumption has been that the "negative birth order" effect on intelligence is a phenomenon searching for an explanation. Explanations that have been developed typically start by citing the Belmont and Marolla data and then referencing other cross-sectional data sets for further support and confirmation. Many theories have been developed to explain these supposed birth order effects. These theories vary in their complexity.

The confluence model (Zajonc & Markus, 1975) is one example that posited a family environmental influence on children's intellectual development. The more adults and older children, the richer the environment. Conversely, with increasing numbers of younger children (e.g., after the birth of each child), the intellectual environment is diluted. Thus, larger families and later birth orders exist in a depressed intellectual environment, which negatively affect intellectual development. The confluence model also contained a "tutoring effect," in which intellectual value was accrued through the tutoring provided to their younger siblings by all non-last-born children; the tutoring effect was added to the confluence model to help explain an IQ discontinuity among last-borns in the Belmont and Marolla data. Blake (1981) adapted the dilution model that had been proposed previously (see, e.g., Kellaghan & MacNamera, 1972; Lasko, 1954; Strodtbeck & Creeland, 1968; Walberg & Marjoribanks, 1976) to help explain the negative birth order finding. Her theory was that more children dilute the parental resources that provide nurture—including support for intellectual development—among developing children.

Recently, however, it has been suggested that the "negative birth order" phenomenon may have been a methodological illusion. Researchers using cross-sectional data—collected from individuals in different families at a specific point in time—have typically assumed that those cross-sectional patterns would match those across siblings within families. Researchers are familiar with the challenges involved when inferring longitudinal change processes from cross-sectional data but have been forced (by lack of available data) to assume that patterns in cross-sectional data would hold when subjected to longitudinal scrutiny. Unfortunately, this assumption does not necessarily hold. Rodgers, Cleveland, van den Oord, and Rowe (2000) compared the patterns from cross-sectional data to those from the few within-family studies that have been run and found them to be entirely different. The negative birth order phenomenon simply disappeared when actual siblings' IQs were compared to one another. They explained this result by noting that in the cross-section, birth order can be an indirect measure of literally thousands of potential biases, including socioeconomic status (SES), maternal health, nutrition, parents' education, parental IQ, quality of schooling, and dozens of other less-obvious processes. If patterns in cross-sectional data and within-family data do not match, then the assumption that birth order in the cross-section is actually measuring differences between siblings in a family, that is, real birth order patterns, is untenable. The finding that hundreds of cross-sectional birth order–IQ patterns (which match one another) do not match the dozen or so within-family patterns (which also tend to match one another) is disturbing. A conservative view would be that the absence of a birth order–IQ relationship in within-family studies suggests that there is no birth order effect on intelligence.

However, existing within-family studies have methodological problems as well, even though their within-family matching partially controls for the thousands of between-family biases that plague the cross-sectional studies. What are these problems? Age matching is one. Researchers have repeatedly expressed concern about age variance because the confluence model predicts different birth order–IQ patterns at different ages (e.g., Zajonc, 2001; Zajonc & Mullally, 1997). Others (e.g., Michalski & Shackelford, 2001; Steelman & Mercy, 1980) have expressed concern about age confounding. Another concern is statistical. Failure to properly separate between-family variance from within-family variance (see, e.g., Jensen, 1980, for discussion) has kept researchers from determining the actual sources of influence on IQ observed in studies of birth order effects. Modeling critical between-family differences that account for large parts of children's variance in intelligence is a methodological innovation that can help identify the location of the actual sources of shared variance between birth order and IQ (if any exists at all). In this article, we will use design and statistical innovations to address these concerns.

In our opinion, there are only two types of past birth order research that shed light on underlying causal processes. The first, already discussed, is research using within-family comparisons. When actual siblings are compared to one another, within-family and between-family sources of variance can be separated. We note, however, that uncontrolled within-family variance is still a problem (although much reduced). For example, as parents age, they typically increase in SES level and also may spend more time at work and less time with their children. Thus, later-born children may, on average, mature in a slightly higher SES environment than their earlier-born siblings, but one in which parents spend less time with them and which therefore may negatively affect their intellectual development. If later-born children have lower IQs, are we observing an effect of being a later-born child (a real birth order effect) or an indirect effect of SES? Obviously, within-family studies are no panacea.

The second type of birth order research includes the few studies that have used design innovations to study the relationship between birth order and IQ. These studies are substantially undervalued in past treatment of the birth order literature. We will carefully examine several of these to illustrate how design innovations can solve many of the problems with past birth order research. Typically, these design innovations require the use of within-family data.

DESIGN INNOVATIONS IN PAST BIRTH ORDER RESEARCH

Most past birth order research has been based on one of two design approaches. In one approach, a cross-section has been obtained, observed on some outcome, and that outcome has been classified by birth order/family size category. In the other approach, individuals with some interesting feature in common (e.g., presidents, strippers, smokers, birth order researchers, etc.; see Rodgers & Thompson, 1985) have been studied to determine if their birth order is the same as that in the general population. Both designs have many deficiencies (see Adams, 1972; Kammeyer, 1967; Rodgers & Thompson, 1985; Schooler, 1972).

In a few cases, innovative and unusual approaches have been used to study birth order. One early study anticipated recent criticisms of birth order literature. Thurstone and Jenkins (1929) obtained IQ information from the Institute for Juvenile Research for more than 10,000 observations, which contained many siblings. After observing trends in the cross-sectional data, they noted that "the summary . . . is defective for our purposes in that several antagonistic factors are there at work without any control" (p. 644). They were concerned about the biases built into cross-sectional data and thus limited their analysis to the study of siblings. Their interpretation of their findings has, unfortunately, gone unheeded by most birth order researchers since then:

> If the intelligence of children is improved by the experience of parents in bringing up children, then it is conceivable that such experience would affect the comparison of first and second born children. . . . It is more probable, however, that the causal relation is more strongly in the reverse direction, namely that . . . the children are bright because the parents are bright. (p. 645)

Koch (1954) followed Thurstone and Jenkins's (1929) lead by using an even more restricted family design. Her study was based on "360 five- and six-year-olds from two-child, intact, native-born, White, urban families" (Koch, 1957, p. 176). She, too, obviously appreciated the value of controlling the selection biases inherent in cross-sectional designs.

Although most past birth order research was based on cross-sectional designs, a few other studies followed the example that Thurstone and Jenkins (1929) and Koch (1954) set by using within-family data. Included among those are Olneck and Bills (1979), Berbaum and Moreland (1980), Pfouts (1980), Galbraith (1982), Rodgers (1984), Retherford and Sewell (1991), and Rodgers et al. (2000). Rodgers (2001) reviewed the consistent non-relationship between birth order and IQ/achievement found in those studies.

McCall (1984) used an innovative approach to test a prediction of the confluence model: the birth of a child should cause a discontinuity in the IQ scores of older siblings. He investigated whether there were IQ fluctuations among older siblings after the birth of a younger sibling. He found evidence for such a fluctuation, one of the few examples in the literature suggestive of a within-family influence in intelligence.

Guo and Van Wey (1999) used difference models to partially control for unobserved heterogeneity (i.e., selection bias) in studying birth order and family size relations to achievement scores. They used two types of these difference models, one a within-individual model and the other a model in which siblings were compared to one another. In cross-sectional versions of their data, they found the usual negative relation between birth order/family size and IQ. When the unobserved factors that came from differences between families were partially controlled through their difference models, the relationship virtually disappeared.

We review these design innovations for two reasons. First, they help frame the logic of what information is informative about what causes birth order/IQ patterns. Obviously, theoretically based research (that derives from predictions emerging from theoretical statements about within- and between-family processes) and within-family data are important features of past research contributing to our understanding. Second, they set the stage for design innovations that we present in the current article. These include the use of longitudinal within-family data that can be used to obtain samples of siblings at fixed ages. In addition, an analytic approach that has not previously been applied to birth order data—multilevel modeling—will be used to explicitly separate within- and between-family sources of variance.

To implement these innovations, we used data that contained two types of information. First, we needed variables with information about individual children's intellectual ability, and second, a family-level variable to help us remove the influence of between-family factors from that of within-family factors affecting intelligence.

These goals were met through the use of data from the National Longitudinal Survey of Youth (NLSY79), a project developed by the U.S. Department of Labor, which contracted with the Center for Human Resource Research (CHRR, 1999) at The Ohio State University to conduct the survey. Initial data collection on a nationally representative household sample of individuals between 14 and 21 years of age began in 1979. Data collection was later expanded to include the children of the 14- to 21-year-old women who were assessed in the original 1979 sample. Beginning in 1986 and continuing biannually through the present, measures have been collected of these children's cognitive ability, temperament, motor and social development, behavior problems, home environment, and of their attitudes toward self and others. These characteristics, along with the large sample size, made this data set ideal for this project. In the original NLSY79 sample, there were 6,283 women. By 1998, a total of 10,918 children had been born to these women and included in the NLSY79 data set. We used data from the NLSY79 from the beginning of the child data collection program in 1986 up to 1998.

METHOD

Sampling Technique

Our sampling technique minimized certain threats to validity that other studies on this topic have faced. An *age-snapshot* technique was used to minimize cohort and maturation influences on criterion variables while maintaining large sample size. We applied this technique to obtain two cohorts, one composed of 7- and 8-year-old children and the other of 13- and 14-year-old children. In the NLSY data set, children were assessed on our chosen criterion variables (described below) every 2 years starting in 1986 and continuing through 1998. We extracted children's age-standardized criterion scores from the total NLSY sample at that assessment time point between 1986 and 1998 when the child was within the target age range. By holding approximately constant the children's age with the snapshot technique, we reduced the threat to validity of differential, child-age related, within-family maturation influences that might covary with the children's different ages. To improve the generalizability of our results, we chose two different target age ranges, consisting of separate samples of 7- and 8-year-olds (84-96 months) and of 13- and 14-year-olds (156-168 months), respectively. All children who were in the target age range (84-96 months or 156-168 months) at any time between 1986 and 1998 were included in each sample. The 12-year length of the sampling window caused some children to be potentially sampled twice, once when they were 7 or 8 and once when they were 13 or 14. However, any child that was sampled twice was placed in the 13- to 14-year-old cohort so that cohorts were nonoverlapping with respect to individual children. Both cohorts were skewed with respect to birth order such that there were relatively few fifth, sixth, and seventh born children and many more first and second born children. Therefore, we analyzed the first to the fourth born children in both cohorts. Table 1 shows the number of children of each birth order in each sample, and Table 2 shows the number of families of different sizes in each cohort.

Our total sample sizes for the two age snapshots were 1,902 and 1,769 for the 7- to 8-year-olds and 13- to 14-year-olds, respectively. The total number of families was 2,972. The number of valid cases available on our different criterion and predictor variables caused sample size to vary slightly across subsequent analyses.

Outcome Variables Investigated

A set of cognitive ability measures was selected to operationalize child's intelligence, along with a control

TABLE 1: Number of Children per Birth Order

	In Each Cohort		
	7-8	*13-14*	*In Total Sample*
First born	808	1,036	1,844
Second born	640	522	1,162
Third born	347	160	507
Fourth born	107	51	158
Total *N*	1,902	1,769	3,671

TABLE 2: Distribution of Number of Families of Different Sizes

	Number of Families in Each Cohort		*Number in Total Sample*
	7-8	*13-14*	
1-child families	1,329	1,008	2,337
2-child families	259	314	573
3-child families	17	43	60
4-child families	1	1	2
Number of families	1,606	1,366	2,972

NOTE: 1-child families refers to families that had one child present in the sample, 2-child families refers to families that had two children present in the sample, and so forth.

variable to permit better disentanglement of between-family and within-family influences on intelligence. We operationalized children's intelligence by using the different sections of the Peabody Individual Achievement Test (PIAT; Dunn & Dunn, 1981; Dunn & Markwardt, 1970). Age-standardized PIAT Mathematics, Reading Recognition, and Reading Comprehension subtest scores were used. For all PIAT subtests, higher scores indicate greater cognitive ability. More detailed information on the PIAT assessments is available from Dunn and Dunn (1981) and Dunn and Markwardt (1970).

Mother's Age at Birth of First Child as a Control Variable

In addition to birth order as the central predictor variable, we chose to investigate the effect of a control variable on the relationship between birth order and children's intelligence. Birth order is correlated with many factors that vary between families. These factors include cultural norms influencing family size, SES, access to birth control information or methods, or a woman's age when she becomes a mother. Maternal age was selected as a control variable to use in our sample.

Mothers' ages can affect both the number of children they have and the conditions under which their children develop. Environments influence age of access to potential mates and can play important roles in influencing mating behavior and the desire to have children. Arguably, environments that facilitate early childbearing do not simultaneously facilitate mothers' education and career success because women who attain a high level of education and/or invest considerable time and effort into a career would seem likely to delay childbearing. These women may be more likely to support the intellectual development of their children and they may be likely to value education and to model behaviors that help children succeed in school. Thus, motherhood age is not only a measure of a mother's likely maturity but is probably also a measure that captures many between-family environmental influences on children's intelligence. We expected that maternal age values would be positively associated with child intelligence.

Multilevel Modeling and Nested Data

The extraction of these variables for each age-snapshot sample led to the creation of a nested, hierarchical data structure where each child was nested within its family (as indicated by the common mother). This nesting allows the examination of both within-family (birth order) and between-family (maternal age) sources of variance on children's intelligence. There are many potential within-family influences, which include birth order, changes in the family's SES over time, or other factors varying from child to child within a family. These within-family influences can be obvious or they can be subtle. For instance, research on physical attractiveness (Eagly, Ashmore, Makhijani, & Longo, 1991; Snyder, Tanke, & Berscheid, 1977) implies that variation in children's attractiveness within the same family may influence parenting behavior. There are almost certainly even more potential between-family influences, including such factors as mother's intelligence, parental disciplinary style, or the quality of schools in a given district; these between-family influences do not vary across children in a family. Although the level of a particular between-family influence may vary across families, different children within a family are not exposed to different levels of these influences. For example, a mother's age when she has her first child remains constant for all subsequent children in her family. Furthermore, we expect this variable to covary with—and to indirectly measure—many other variables, as discussed above.

Researchers have known for some time that it can be misleading to analyze group-level data (e.g., family-level data) and use these results to draw inferences about individuals within the groups (e.g., Robinson, 1950). Kreft, de Leeuw, and Aiken (1995) considered a case where data were analyzed at the level of individual workers and educational level was positively associated with income, but when these data were analyzed at the higher level of industry, the relationship reversed. Previous studies of birth order and intelligence have faced just this problem. Birth order is an individual-level variable and individuals are nested within families. However, birth order has often been analyzed in an aggregated fashion, ignoring the child's family membership. The consequence is that the influences of family level variables (such as how old mothers are when they begin to have children) on the outcome measure are not partialed out of any within-family birth order effect. To determine any true effects of birth order, such between-family influences must be controlled. In addition to the potential invalidity of analyses that do not take between-family influences into account when within-family effects are being estimated, analyses of this type violate the assumptions of traditional regression analysis (see, e.g., Barcikowski, 1981; Moulton, 1986; Scariano & Davenport, 1987; Scott & Holt, 1982; Walsh, 1947). Specifically, analysis of individual-level data, when those individuals are clustered within families, violates the assumption of independence of residuals. That is, individuals within families tend to be more alike than do individuals in different families, and thus, these observations (and their associated error terms) are not independent. Such a violation can cause considerable bias in estimates of effects and their standard errors. It is entirely possible that effects reported in previous studies that have used traditional regression techniques for analysis may owe their significance to the use of an inappropriate statistical model that does not appropriately partition the sources of variance present in nested data.

We therefore used multilevel models to investigate the association between birth order and intelligence (Goldstein, 1995; Kreft & de Leeuw, 1998; Raudenbusch & Bryk, 2002). Unlike traditional regression analyses, multilevel models allow hierarchical partitioning of variance and reduce the magnitude of the above threats to statistical conclusion validity. With data such as ours, these models yield more accurate standard error estimates, leading to more accurate significance tests. In addition, multilevel models allow simultaneous estimation of group- and individual-level parameters as well as estimation of cross-level interaction terms, where the effect of a between-family variable (such as mother's intelligence) on a within-family variable (child's intelligence) can be more accurately modeled without the consequences that follow from violating the assumptions of traditional regression.

Description of Multilevel Model Sequence

We coded birth order using dummy variables and fit three models of increasing complexity to each of the children's PIAT subtest scores. The use of dummy coding for birth order allowed for representation of even nonlinear relationships between birth order position and outcome variables. The first model estimated the basic birth order effect, ignoring cohort and between-family influences on intelligence that could be encapsulated by mother's age at birth of first child. The second model included cohort and the third model also included maternal age as a predictor. In the following, we present our model specifications using standard notation (cf. Nezlek, 2001). Parameter estimates are reported in the Results section.

Multilevel models can be fit using a variety of software packages. These programs vary in their flexibility, options, and computational robustness but they all effectively implement multilevel random coefficient modeling (MRCM). MLwiN, LISREL's multilevel modeling module, HLM, and PROC MIXED in SAS are all common programs used to fit these models. The rapid pace of software development makes it difficult to maintain a current or complete listing but there are also many other programs that effectively fit multilevel models. For our analyses, we used LISREL's multilevel modeling module (see Jöreskog, Sörbom, DuToit, & DuToit, 2001).

Model 1: The birth order effect. The main purpose of this model was to obtain initial estimates of the IQ-related outcome values for each birth order. Our basic model expressed each outcome variable, y_{ij}, as a function of birth order represented as a set of four dummy variables ($d_{ij(k)}$), where i represents child, j represents family, and k represents birth order with values 1, 2, 3, and 4. For a child of birth order k, the value of the corresponding dummy variable $d_{ij(k)} = 1$, and the values of the other three dummy variables would be 0. The model also included a child-level (Level 1 [L1]) residual term, $r_{ij(k)}$, associated with each birth order, along with a family-level (Level 2 [L2]) residual term u_j. This model is given by

$$y_{ij} = \sum_k \beta_k d_{ij(k)} + \sum_k r_{ij(k)} + u_j . \tag{1}$$

When this model is fit to data, results include estimates of the four β_k terms that correspond to the mean level of the outcome variable at each birth order, the between-child residual variance in the outcome variable at each birth order, and the between family residual variance. This model specification avoids having to assume that the residual variance in the outcome variable is the same for each birth order.

The presence of a birth order effect would be reflected by significant differences among the four β_k coefficients. For this and other models, we will report the result of a test of the null hypothesis that the four β_k coefficients are equal. Rejection of this null hypothesis implies a significant effect of birth order on the outcome variable in question.

Model 2: Adding cohort as a predictor. The second model served to determine whether the pattern of birth order differences was consistent across the two age cohorts. Cohort was coded as a dummy variable, C_{ij} (0 for the younger sample, 1 for the older sample) and was defined as an L1 (child level) variable because each child is a member of one cohort or the other and no child is in both cohorts. Essentially, cohort functions as a binary measure of age. (Note that cohort could not be considered as an L2 [family-level] variable because some families included children in each cohort.) This data structure and the research question imply that we must investigate the interactive effect of cohort and birth order on each outcome variable. To that end, we constructed four product variables, each a product of C_{ij} and one of the birth order dummy variables, and we included these product variables in the model. This extended model is given by

$$y_{ij} = \sum_k \beta_k d_{ij(k)} + \sum_k \beta_{C(k)} C_{ij} d_{ij(k)} + \sum_k r_{ij(k)} + u_j. \quad (2)$$

Of primary interest in evaluating results from this second model will be whether the birth order effect itself, as represented by differences among the four β_k coefficients, changes from the first model. In addition the coefficients $\beta_{C(k)}$ will indicate cohort differences in mean level of the outcome variable for each birth order.

Model 3: Adding mother's age at birth of first child as a predictor. In our third and final model, we included maternal age at birth of first child (centered at the grand mean), denoted M_j, as an L2 predictor of birth order effects. Although birth order itself varies within family, maternal age at first birth varies only between family. If adding to the model a variable that varies only between family reduces or eliminates the birth order effect, this would provide strong evidence that supposed birth order effects are really due to factors outside the family that happen to correlate with birth order, not due to the effects of birth order itself. To investigate the effects of maternal age, we introduced cross-level interactions of grand-mean-centered maternal age and birth order. This was accomplished via the inclusion of effects of four product variables, each a product of maternal age (M_j), and one of the four dummy variables representing birth order. The addition of these four additional effects produce the final extension of the previous models:

$$y_{ij} = \sum_k \beta_k d_{ij(k)} + \sum_k \beta_{C(k)} C_{ij} d_{ij(k)} + \sum_k \gamma_k \left(M_j d_{ij(k)} \right) + \sum_k r_{ij(k)} + u_j. \quad (3)$$

The primary question in examining results from this model will involve whether a birth order effect is still present after introducing this family-level predictor. That is, are the four β_k significantly different? We used this significance test, as well as visual inspection of the individual parameter estimates for each of the birth order coefficients, to investigate our hypothesis. Because our models are nested in the sequence presented above, we were able to use χ^2 difference tests on the models' $-2\ln L$ values to compare the fit of models to each other as we progressed through the three-model sequence.

RESULTS

The series of three models described in the previous section was applied to data from our sample as described earlier. For each sample, three different PIAT subtests were used as outcome variables (Math, Reading Recognition, and Reading Comprehension). For all outcome variables, χ^2 difference tests conducted to compare each model to the subsequent model indicated that fit significantly improved from model to model.

Model 1 Results

Our first model showed significant birth order effects for all PIAT outcome variables. This significance was indicated both by visual inspection and by the test of the null hypothesis that the four β_k coefficients were equal. Table 3 shows parameter estimates for each birth order ($\beta_{[k]}$) as well as the four L1 residual variances (Var[$r_{ij(k)}$]) and the L2 residual variance (Var[u_j]). In this and the remaining tables, values in parentheses are estimated standard errors associated with each parameter estimate. An approximate significance test of each parameter estimate can be conducted by dividing the parameter estimate by its standard error. If the result is 2 or larger, the estimate is significant. In Table 3, all parameter estimates were statistically significant, as was the test of the birth order effect. The pattern of coefficients for the dummy variables also reveals a decline in the mean level of each outcome variable in later birth orders, an apparent negative effect of birth order on measures of intelligence.

Model 2 Results

Results for Model 2 are provided in Table 4. The addition of cohort as an L1 predictor did not alter the statistical significance or the pattern of the birth order effect on each outcome variable. However, the significant coefficients for the product variables representing the interaction of cohort and birth order indicate that there were significant cohort differences in mean levels of the outcome variables at each birth order. Of interest, the results indicate that compared to the norming sample used for age standardization of the PIAT scores, the 13- to 14-year-old cohort (coded 1) was slightly less intelligent than the 7- to 8-year-old cohort. This finding in no way affects the conclusion of this analysis: Birth order remains a significant predictor of intelligence when cohort is included as a predictor. Considering residual variances, results for Model 2 show a slight decrease in

TABLE 3: Model 1: Initial Estimates of Fixed Birth Order Effects

	Math	*Reading Recognition*	*Reading Comprehension*
	$N_{families}$ = *2,475;* $N_{children}$ = *3,641*	$N_{families}$ = *3,370;* $N_{children}$ = *3,681*	$N_{families}$ = *2,370;* $N_{children}$ = *3,428*
Birth order effects			
β_1	99.47 (0.29)	103.64 (0.33)	100.62 (0.31)
β_2	98.48 (0.35)	101.91 (0.40)	99.96 (0.38)
β_3	97.48 (0.52)	99.49 (0.59)	98.91 (0.57)
β_4	97.19 (0.95)	98.90 (0.94)	97.54 (0.96)
L1 residuals			
Var ($r_{ij[1]}$)	107.34 (5.75)	137.89 (7.45)	109.05 (6.32)
Var ($r_{ij[2]}$)	97.12 (6.08)	125.52 (7.89)	105.84 (6.93)
Var ($r_{ij[3]}$)	102.36 (8.65)	121.15 (10.62)	107.04 (9.78)
Var ($r_{ij[4]}$)	111.43 (16.02)	95.79 (15.40)	96.38 (15.47)
L2 residual			
Var (u_j)	50.18 (4.44)	71.71 (5.78)	61.82 (5.11)
–2 lnL value	28518.62	29416.61	27146.45
Test of birth order effect	$\chi^2(1) = 6.67$	$\chi^2(1) = 30.33$	$\chi^2(1) = 11.39$
	$p = .010$	$p = .000$	$p = .001$

NOTE: L1 = Level 1; L2 = Level 2. Standard errors are in parentheses.

TABLE 4: Model 2: Effect of Cohort as an Additional Predictor of IQ

	Math	*Reading Recognition*	*Reading Comprehension*
	$N_{families}$ = *2,475;* $N_{children}$ = *3,641*	$N_{families}$ = *2,464;* $N_{children}$ = *3,635*	$N_{families}$ = *2,370;* $N_{children}$ = *3,428*
Birth order effects			
β_1	101.39 (0.43)	105.58 (0.50)	106.19 (0.45)
β_2	99.71 (0.46)	103.55 (0.53)	104.55 (0.48)
β_3	98.44 (0.62)	100.71 (0.70)	101.80 (0.65)
β_4	97.80 (1.14)	99.58 (1.13)	100.10 (1.06)
Cohort effects			
$\beta_{C(1)}$	–3.39 (0.57)	–3.41 (0.66)	–9.43 (0.58)
$\beta_{C(2)}$	–2.89 (0.68)	–3.80 (0.78)	–10.15 (0.69)
$\beta_{C(3)}$	–3.77 (1.10)	–4.78 (1.22)	–10.52 (1.10)
$\beta_{C(4)}$	–3.21 (2.02)	–3.96 (1.97)	–11.15 (1.79)
L1 residuals			
Var ($r_{ij[1]}$)	106.34 (5.67)	138.53 (7.38)	99.45 (5.53)
Var ($r_{ij[2]}$)	96.45 (6.01)	125.96 (7.82)	92.50 (5.98)
Var ($r_{ij[3]}$)	100.00 (8.45)	118.99 (10.39)	92.79 (8.37)
Var ($r_{ij[4]}$)	109.38 (15.70)	93.99 (15.03)	76.64 (12.43)
L2 residual			
Var (u_j)	48.47 (4.36)	67.35 (5.65)	48.23 (4.32)
–2 lnL value	28455.19	29351.87	26610.78
Test of birth order effect	$\chi^2(1) = 10.32$	$\chi^2(1) = 30.12$	$\chi^2(1) = 35.03$
	$p = .001$	$p = .000$	$p = .000$

NOTE: L1 = Level 1; L2 = Level 2. Standard errors are in parentheses.

the L1 residual variances for each outcome variable as compared to Model 1 and also slight decreases in the L2 residual for each PIAT outcome variable.

Model 3 Results

Our third model provided the critical test of our hypothesis. If the birth order effect is primarily a within-family (L1) phenomenon, then accounting for between-family (L2) variance should have little influence on the effect. However, if the birth order effect is primarily due to factors varying between families, then the addition of a *between*-family variable could reduce or even eliminate the effect. We included maternal age as an L2 predictor to test this idea. Table 5 shows that when maternal age is included, the test of the birth order effect for all three PIAT outcome variables becomes nonsignificant. The negative birth order effect was completely eliminated for PIAT math scores ($p = .810$). An apparent small, negative birth order effect remained but was reduced to nonsignificance for PIAT reading recognition ($p = .157$) and reading comprehension ($p = .077$). It appears that virtually all of the birth order effect may be due to uncon-

TABLE 5: Model 3: Effect of a Between-Family (L2) Control Variable on the Birth Order Effect

	Math	*Reading Recognition*	*Reading Comprehension*
	$N_{families}$ = *2,207;* $N_{children}$ = *3,313*	$N_{families}$ = *2,198;* $N_{children}$ = *3,308*	$N_{families}$ = *2,121;* $N_{children}$ = *3,127*
Birth order effects			
β_1	99.11 (0.56)	103.17 (0.65)	104.35 (0.57)
β_2	98.95 (0.51)	102.60 (0.59)	103.81 (0.53)
β_3	99.03 (0.67)	101.84 (0.75)	102.72 (0.70)
β_4	99.46 (1.45)	100.61 (1.41)	102.19 (1.28)
Cohort effects			
$\beta_{C(1)}$	–0.39 (0.72)	–0.25 (0.83)	–6.87 (0.72)
$\beta_{C(2)}$	–0.44 (0.79)	–1.15 (0.91)	–8.05 (0.81)
$\beta_{C(3)}$	–3.56 (1.20)	–4.39 (1.33)	–10.76 (1.22)
$\beta_{C(4)}$	–2.66 (2.12)	–3.61 (2.05)	–10.16 (1.80)
Maternal age effects			
$\gamma_{M(1)}$	0.74 (0.09)	0.79 (0.10)	0.62 (0.09)
$\gamma_{M(2)}$	0.85 (0.11)	0.92 (0.13)	0.72 (0.12)
$\gamma_{M(3)}$	0.44 (0.19)	0.72 (0.21)	0.40 (0.20)
$\gamma_{M(4)}$	0.83 (0.39)	0.59 (0.37)	0.95 (0.34)
L1 residuals			
Var ($r_{ij[1]}$)	109.55 (5.91)	142.30 (7.70)	100.82 (5.69)
Var ($r_{ij[2]}$)	95.92 (6.11)	125.32 (7.98)	93.48 (6.11)
Var ($r_{ij[3]}$)	102.84 (8.80)	121.85 (10.82)	96.96 (8.79)
Var ($r_{ij[4]}$)	111.46 (16.04)	94.08 (15.16)	69.85 (11.68)
L2 residual			
Var (u_j)	42.89 (4.36)	61.36 (5.68)	42.54 (4.30)
–2 lnL value	25825.41	26654.00	24189.66
Test of birth order effect	$\chi^2(1) = 0.058$	$\chi^2(1) = 2.01$	$\chi^2(1) = 3.11$
	$p = .810$	$p = .157$	$p = .077$

NOTE: L1 = Level 1; L2 = Level 2. Standard errors are in parentheses.

trolled between-family variance and that when this between-family variance is controlled, birth order is not a significant determinant of children's intelligence. The reduced birth order effect also is reflected by the smaller differences between the β_k coefficients for each outcome variable.

As expected, the addition of maternal age as an L2 predictor reduced the L2 residual variance for all outcome variables but did not reduce the L1 residual variances. This follows from the fact that maternal age does not vary within family.

DISCUSSION

We begin our discussion by summarizing the logic of our design and reviewing our conclusions within the context of that logic. Next, we broaden our conclusions to discuss future directions for research on family structure and intelligence.

The design logic is straightforward. If IQ/achievement score means decline across birth order, the cause of those declines may lie either within the family or outside of the family. Researchers have consistently interpreted those causes to lie within the family and have built within-family models to explain the declines. We argue that this interpretation is empirically unjustified. In this study, we have controlled for age and for the cross-sectional confound with two design innovations and with a unique analytic strategy. First, we used within-family data, which virtually all birth order theorists have recognized as having critical advantages. Second, we compared siblings to one another at fixed ages (as opposed to the usual approach of using a fixed time). The ages we chose span the cross-over period suggested by Zajonc and Mullally (1997) as occurring somewhere around age 12. Our analytic approach used multilevel modeling, which explicitly partitioned the variability in our PIAT measures into within-family variance and between-family variance.

The NLSY data could hardly speak more clearly and conclusively within the context of this particular design and analytic strategy. When we controlled for a single variable (maternal age) reflecting only between-family variance, previously observed birth order effects disappeared or became nonsignificant. This result suggests that the fundamental cause of supposed birth order effects lies between, not within, families. This finding was observed in data that originated as a national probability sample. (We note, however, that it was the mothers who were randomly sampled through a sample of U.S. households in 1979 to reflect the whole U.S. population of adolescents in that year; furthermore, attrition, nonresponse, and selection for childbearing are additional threats to the external validity of this sample.)

This study is fundamentally different from previous work but supports a conclusion for which there is mounting evidence. In addition to its multilevel modeling approach, it uses a unique age-snapshot sampling technique. For instance, in past research, Rodgers et al. (2000) used scores based on a time-matching design rather than an age-matching design (actually, they averaged scores throughout a 2-year period at two different points in time and then also analyzed age-restricted subsets of each of those fixed time periods). Their logic was based on the presumption that if past patterns in cross-sectional data did actually originate primarily within the family, then those patterns should be approximately the same using within-family and between-family data. They reviewed a number of within-family data sources showing that this was not the case and analyzed an overlapping but different subset of the NLSY data used in the current study to show that the birth order PIAT pattern was flat within families. Armor (2001) and Rodgers, Cleveland, van den Oord, and Rowe (2001) extended this result by also considering both Peabody Picture Vocabulary (PPVT) and Digit Span patterns. These findings are consistent with ours. The sources of the often-found birth order–IQ relationship appear to lie outside the family. Some researchers have argued that the within-family processes really are there but are hidden by age confounds (e.g., Zajonc, Markus, & Markus, 1979; Zajonc & Mullally, 1997). Rodgers et al. (2000) addressed this issue by looking at patterns on both sides of the age cross-over identified by Zajonc and Mullally. Although this design eliminated time confounds as a source of bias, it still did not address concerns regarding the differential ages of the siblings compared to one another within the families (e.g., Zajonc, 2001). The design in the current study explicitly addresses this concern by age matching rather than time matching. Its findings, once again, suggest that birth order does not affect intelligence. Of course, age matching creates some potential threats to internal validity as well, caused by period effects, for example. But when results from both age matching and time matching designs converge on the same conclusion, these threats are less plausible. It is well known to demographers and developmental researchers that age, cohort, and period effects are confounded in such a way that only two of those have independent influences (e.g., Schaie, 1965). Implicitly, then, cohort confounds are accounted for by combining results from the two studies controlling for age and period effects.

What direction should researchers take from here? First, we should accept that previous claims of a consistent and/or potent birth order effect on IQ/achievement in cross-sectional studies were misattributions. When we look inside families at a fixed time period to find these patterns, they are typically absent (Rodgers et al., 2000). When we looked inside ages in the current study, these patterns initially appeared to be there, consistent with past research findings. However, we showed that these patterns actually originated in the differences between families and not through processes operating within the family. Obviously, new studies with similar designs—especially those based on within-family data—need to be run to determine whether these results replicate. So far, there is considerable consistency across the several within-family studies reviewed in the introduction to this article. Second, if birth order/IQ patterns are only apparent, and if the causal factors at work are located outside the family, then new emphasis on such variables as maternal age at first birth could be fruitful. Third, researchers might go beyond the simple concept of maternal age as encapsulating diffuse between-family influences and look at how more specific between-family factors influence children's intelligence.

Our data suggest that instead of constructing environments that differentially influence intellectual development across children, parents and families support intellectual development similarly for all children within a given family. Rather than being due to differences in birth order, the substantial differences among children in intellectual ability and achievement at a given age or at a given time period (age-adjusted) more likely derive from differences between families. Parental IQ (passed on through both genetic and environmental mechanisms), parental education, SES differences, neighborhood effects, and school effects are all examples of additional between-family influences that are likely contenders to explain substantial variance in childhood and adolescent IQ/achievement. In the context of research on birth order effects on intelligence, these are all confounds. Birth order is confounded with some of these true between-family influences on intelligence, and the current study suggests that such confounds are responsible for its apparent effects.

REFERENCES

Adams, B. N. (1972). Birth order: A critical review. *Sociometry*, *35*, 411-439.

Anastasi, A. (1956). Intelligence and family size. *Psychological Bulletin*, *53*, 187-209.

Armor, D. J. (2001). On family size and intelligence. *American Psychologist*, *56*, 521-522.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, *6*, 267-285.

Belmont, L., & Marolla, F. A. (1973). Birth order, family size, and intelligence. *Science*, *1982*, 1096-1101.

Berbaum, M. L., & Moreland, R. L. (1980). Intellectual development within the family: A new application of the confluence model. *Developmental Psychology*, *16*, 506-515.

Blake, J. (1981). Family size and the quality of children. *Demography, 18*, 421-442.
Center for Human Resource Research. (1999). *The 1998 NLSY79 child assessments: Selected tables.* Columbus: The Ohio State University, CHRR User Services.
Dunn, L. M., & Dunn, L. M. (1981). *PPVT-revised manual.* Circle Pines, MN: American Guidance Service.
Dunn, L. M., & Markwardt, F. C., Jr. (1970). *Peabody individual achievement test manual.* Circle Pines, MN: American Guidance Service.
Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 90*, 1-20.
Ernst, C., & Angst, J. (1983). *Birth order: Its influence on personality.* New York: Springer-Verlag.
Galbraith, R. C. (1982). Sibling spacing and intellectual development: A closer look at the confluence model. *Developmental Psychology, 18*, 151-173.
Galton, F. (1874). *English men of science.* London: McMillan.
Goldstein, H. (1995). *Multilevel statistical models.* London: Edward Arnold.
Guo, G., & Van Wey, L. K. (1999). Sibship size and intellectual development: Is the relationship causal? *American Sociological Review, 64*, 169-187.
Hall, E. (1986, February). Mining new gold from old research: Conversation, Robert B. Zajonc. *Psychology Today*, pp. 47-51.
Jensen, A. R. (1980). Uses of sibling data in educational and psychological research. *American Educational Research Journal, 17*, 153-170.
Jöreskog, K. G., Sörbom, D., DuToit, S., & DuToit, M. (2001). *LISREL 8: New statistical features* (2nd ed.). Chicago: Scientific Software International.
Kammeyer, K. I. (1967). Birth order as a research variable. *Social Forces, 46*, 71-80.
Kellaghan, T., & MacNamera, J. (1972). Family correlates of verbal reasoning ability. *Developmental Psychology, 7*, 49-52.
Koch, H. L. (1954). The relation of primary mental ability in five- and six-year-olds to sex of child and characteristics of his siblings. *Child Development, 25*, 209-223.
Koch, H. L. (1957). The relation in young children between characteristics of their playmates and certain attributes of their siblings. *Child Development, 28*, 175-201.
Kreft, I. G. G., & de Leeuw, J., (1998). *Introducing multilevel modeling.* London: Sage.
Kreft, I. G. G., de Leeuw, J., & Aiken, L. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*, 1-22.
Lasko, J. K. (1954). Parent behavior toward first and second children. *Genetics Psychology Monographs, 49*, 97-187.
McCall, R. B. (1984). Developmental changes in mental performance: The effect of a birth of a sibling. *Child Development, 55*, 1317-1321.
Michalski, R. L., & Shackelford, T. K. (2001). Methodology, birth order, intelligence, and personality. *American Psychologist, 56*, 520-521.
Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics, 32*, 385-397.
Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin, 27*, 771-785.
Olneck, M. R., & Bills, D. (1979). Family configuration and achievement: Effects of birth order and family size in a sample of brothers. *Social Psychology Quarterly, 42*, 135-148.
Outhit, M. C. (1933). A study of the resemblance of parents and children in general intelligence. *Archives of Psychology, 149*, 1-60.
Pfouts, J. H. (1980). Birth order, age-spacing, IQ differences, and family relations. *Journal of Marriage and the Family, 42*, 517-531.
Raudenbusch, S., & Bryk, A. (2002). *Hierarchical Linear Models applications and data analysis methods.* Thousand Oaks, CA: Sage.
Retherford, R. D., & Sewell, W. H. (1991). Birth order and intelligence: Further tests of the confluence model. *American Sociological Review, 56*, 141-158.
Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *Sociological Review, 15*, 351-357.
Rodgers, J. L. (1984). Confluence effects: Not here, not now! *Developmental Psychology, 20*, 321-331.
Rodgers, J. L. (2001). What causes birth order-intelligence patterns? The admixture hypothesis, revived. *American Psychologist, 56*, 505-510.
Rodgers, J. L., Cleveland, H. H., van den Oord, E., & Rowe, D. C. (2000). Resolving the debate over birth order, family size, and intelligence. *American Psychologist, 55*, 599-612.
Rodgers, J. L., Cleveland, H. H., van den Oord, E., & Rowe, D. C. (2001). Birth order and intelligence: Together again for the last time? *American Psychologist, 56*, 523-524.
Rodgers, J. L., & Thompson, V. D. (1985). Toward a general framework of family structure: A review of theory-based empirical research. *Population and Environment, 8*, 143-172.
Scariano, S. M., & Davenport, J. M. (1987). The effects of violations of independence assumptions in one-way ANOVA. *American Statistician, 41*, 123-129.
Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64*, 92-107.
Schooler, C. (1972). Birth order effects: Not here, not now! *Psychological Bulletin, 78*, 161-175.
Schvaneveldt, J. D., & Ihinger, M. (1979): Sibling relationships in the family. In W. R. Burr, R. Hill, F. I. Nye, & I. L. Reiss (Eds.), *Contemporary theories about the family: Vol. 1. Research-based theories* (pp. 453-467). New York: Free Press.
Scott, A. J., & Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association, 77*, 848-854.
Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*, 656-666.
Steelman, L. C., & Mercy, J. A. (1980). Unconfounding the confluence model: A test of sibship size and birth order effects on intelligence. *American Sociological Review, 45*, 571-583.
Strodtbeck, F. L., & Creeland, P. G. (1968). The interaction linkage between family size, intelligence, and sex-role identity. *Journal of Marriage and the Family, 30*, 301-307.
Sulloway, F. J. (1996). *Born to rebel.* New York: Vintage.
Thurstone, L. L., & Jenkins, R. L. (1929). Birth order and intelligence. *Journal of Educational Psychology, 20*, 641-651.
Walberg, H. J., & Marjoribanks, K. (1976). Family environment and cognitive development: Twelve analytic models. *Review of Educational Research, 46*, 527-551.
Walsh, J. E. (1947). Concerning the effect of the intraclass correlation on certain significance tests. *Annals of Mathematical Statistics, 18*, 88-96.
Zajonc, R. B. (2001). The family dynamics of intellectual development. *American Psychologist, 56*, 490-496.
Zajonc, R. B., & Markus, G. B. (1975). Birth order and intellectual development. *Psychological Review, 82*, 74-88.
Zajonc, R. B., Markus, H., & Markus, G. B. (1979). The birth order puzzle. *Journal of Personality and Social Psychology, 37*, 1325-1341.
Zajonc, R. B., & Mullally, P. R. (1997). Birth order: Reconciling conflicting effects. *American Psychologist, 52*, 685-699.

Received February 18, 2004
Revision accepted May 13, 2005