When Paths Diverge:

"Errors of Prediction" From Preschool Test Scores to Later Cognitive and Academic Measures

Philip S. Dale, *University of Missouri–Columbia*
Paulette E. Mills, *Washington State University–Pullman*
Kevin N. Cole, *Washington Research Institute*
Joseph R. Jenkins, *University of Washington–Seattle*

Long-term follow-up information on children who have participated in early childhood special education (ECSE) has seldom been available. In the present study, the cognitive and academic performance of 171 thirteen-year-old graduates of 2 ECSE curricula is examined. Although preschool cognitive measures continued to predict later performance significantly, there were numerous examples of "errors of prediction," that is, children whose later academic performance was substantially better or worse than predicted on the basis of a regression equation. Six individual and family factors were individually at least moderately associated with outcome status and collectively strongly correlated with status as "improver" or "decreaser": positive temperament, European American ethnicity, male gender, middle income, initial referral for gross motor concerns, and nonreferral for social concerns.

How much does the past predict the future in human development? In particular, how powerful and enduring is the impact of early adverse experience, including poverty, illness, abuse or neglect, or poor early development? These questions can only be addressed with longitudinal studies, because individuals differ not only in their level of performance at any given point in time but also in the shape of their developmental trajectory. Conventional standardized test scores and statistics generally assume a linear growth model, but nonlinear growth patterns in mental development are commonly observed in longitudinal growth studies (e.g., McCall, Appelbaum, & Hogarty, 1973). In this article, we address one aspect of nonlinear trajectory: children whose later performance is either better or worse than would have been predicted based on early performance.

The great majority of research on long-term outcomes for children with early adverse experience has been conducted with typically developing children identified as being at risk for biological (e.g., prematurity) or environmental (e.g., poverty) reasons. Much less is known about long-term outcomes for children identified on the basis of poor performance in early childhood. Several sets of results suggest that focusing on the impact of child, family, and educational variables on these children would be highly illuminating. First, evidence from studies of compensatory education, for children growing up in poverty, suggest that long-term effects can occur, and they may be different from short-term outcomes (Lazar & Darlington, 1982). Second, research on risk factors in development suggests that no one environmental or organismic factor has a strongly determining role in later outcome but the presence of multiple risk factors greatly increases the probability of an adverse outcome. For example, Rutter et al. (1975) found that the probability of a psychiatric disorder in childhood was strongly correlated with the number of the following risk factors present: family discord, parental criminal or psychiatric involvement, social disadvantage, and poor school environment. Third, there is evidence of substantial individual differences in children's vulnerability to risk factors, so much so that the term *resilience* is often invoked. Werner and Smith (1982) noted that even among their sample of toddlers with substantial risk due to poverty, low parental education, prematurity, or other factors, the rate of developmental difficulties was greatly reduced by the presence of a number of positive circumstances, such as availability of alternate caregivers, at least 2 years' time between children, a workload for the mother that was not excessive, a cohesive family, and a multigenerational network of kin and friends during adolescence. Other research (see Masten & Coatsworth, 1998, for a review) points to the role of an easy temperament; an outgoing, social personality; and a sense of humor as positive factors. Some of the factors on Werner and Smith's list may well be facilitated by such characteristics in children.

Address: Philip S. Dale, Department of Communication Science and Disorders, 303 Lewis Hall, University of Missouri–Columbia, Columbia, MO 65211; e-mail: dalep@health.missouri.edu

Research on risk and protective factors in children at risk, together with the smaller body of research on children with demonstrated poor cognitive and linguistic skills, suggests that the nature of those factors varies with the developmental level of the child and the nature of the poor early performance (Masten & Coatsworth, 1998; Rutter, 1990; Werner, 1999). For example, in very early childhood, a close relationship with a caring parent or other caregiver is important, whereas in later childhood, a positive, outgoing personality and good relationships with peers play an increasingly important role. Later still, individual special talents—what Brooks (1999) called "islands of competence"—may have a substantial effect (Masten & Coatsworth, 1998; see also Werner, 1999). Werner observed that the most important predictive factors for children with learning disabilities and attention-deficit/hyperactivity disorder differed by gender: For girls, low birthweight and family socioeconomic status (standard of education, maternal education) were the most important, whereas for boys, perinatal stress and high activity level, along with family stability, were important at 1 year of age. In contrast, for children with mild mental retardation, the most important predictors, beyond the severity of the initial delay, were family stability and early caregiving environment during the first decade of life, but personality and motivational factors (being upbeat, prompt, and staying on task) became much more important after that (Werner, 1999). Thus, it would be premature to assume that the findings on protective factors based on children at risk, or at one particular developmental level, will generalize to children at other levels or to those identified by poor early performance.

The Longitudinal Comparison Project is a uniquely long-term follow-up study of a group of children who participated in an early childhood special education program as preschoolers. They were randomly assigned to one of two different programs, Direct Instruction or Mediated Learning, for 1 or more years as preschoolers and have been followed with annual evaluations since that time. Most are now in their late teens. The overall design and results of this project are reported in Dale and Cole (1988); Cole, Dale, Mills, and Jenkins (1993); and Mills, Dale, Cole, and Jenkins (1995). Very few main effects of the program have been observed, but there has been a continuing series of significant aptitude-by-treatment interactions. Students who initially performed at a lower level have tended to do better if they were in the Mediated Learning program, whereas students who initially performed at a higher level have tended to better if they were in the Direct Instruction program.

The starting point for the present investigation was the continuing pattern of substantial correlations, from early cognitive measures to later cognitive and academic measures, despite differences in program and life histories. For example, the multiple correlation for the prediction of academic performance at age 13 from preschool entry scores for the *McCarthy Scales of Children's Abilities* General Cognitive Index (McCarthy, 1972) and the *Peabody Picture Vocabulary Test–Revised* (PPTV-R; Dunn & Dunn, 1981) was .70. A correlation of this magnitude could be indicative of the intervention's strong but constant effect. However, the fact that it holds across a sample of two groups that experienced quite different interventions and were composed of individuals who had enormously diverse lives after preschool strongly suggests a different and somewhat discouraging conclusion—that environment has had only a modest effect. However, a correlation of .70 is far from perfect. We believe that the high correlation in fact makes the study of individuals who diverge from the projected path all the more interesting. In this study, we identified a group of students who diverged in either a positive or a negative direction and then looked at a number of possible organismic, family, and school influences on later academic achievement. We also explored the relationship between these divergences in academic development and cognitive development as assessed by intelligence tests and school placement experience.

Method

Only a summary of the intervention phase and follow-up phase of this project is provided here; more complete information is provided in Dale and Cole (1988), Cole et al. (1993), and Mills et al. (1995).

Participants: Intervention Sample

Over a 4-year period, 206 children were provided early intervention services in a laboratory school located in the United States. Students were eligible for special education services according to state guidelines. Upon entry in the program, their mean age was 4.9 years and their mean IQ was 76.7, as measured by the *McCarthy Scales of Children's Abilities*. The participants were 142 boys and 64 girls. The sample was composed of 120 European American children; 66 African American children; and 20 children who were Hispanic, Pacific Islander, Asian, Native American, or other. In the state in which the study was conducted, children qualify for special education if they exhibit a delay of at least 1.5 standard deviations on a normed measure in two or more of five areas (language, gross motor, fine motor, cognitive, or social-emotional development) or if they exhibit a delay of at least two standard deviations below the mean in one of these areas.

Classes and Assignment for Intervention

Preschool participants (entering ages between 3 years and 5 years 11 months) attended class 2 hours per day, 5 days per week, for 180 school days. There were six preschool classes per year, three for each of the two programs, with 12 students in each class. One of the three classes for each program con-

tained four typically developing students and eight children with disabilities. The other two classes for each program included only students with disabilities. Kindergarten children attended class 5.5 hours per day, 5 days per week, for 180 school days. There was one such class for each program, with 14 students per class. Students were randomly assigned to a program, and to a classroom within the program.

During the intervention phase of this project, children participated in Direct Instruction (DI) or Mediated Learning (ML) programs. The DI program was based on the educational approach of Englemann (1980). Instruction was explicit, teacher-directed, and fast-paced, with specific procedures for error correction, all directed toward very detailed academic learning outcomes. The specific curricula used for this program were *Distar Language I and II* (Englemann & Osborne, 1976), *Distar Arithmetic I and II* (Engelmann & Carnine, 1975), and *Distar Reading I and II* (Engelmann & Bruner, 1974). The ML program (Osborne & Sherwood, 1984), derived from the work of Feuerstein (Feuerstein, Rand, Hoffman, & Miller, 1980), emphasized cognitive rather than academic learning. Children were taught a variety of thinking strategies, including planning ahead, distinguishing the relevance of information, generating multiple solutions, and evaluating their performance. To promote generalization, teachers followed the child's lead rather than presenting prescribed materials, elicted responses rather than modeling for imitation, and dispersed instruction throughout the day rather than confining topics to time blocks. Further information about these programs and about fidelity of implementation is provided in Cole et al. (1993) and Notari-Syverson, Cole, Osborne, and Sherwood (1996).

Intervention Phase Measures

Children participated in the intervention phase for 1 or more years. In the present report, the following subset of measures obtained during each child's 1st year were used. Pretests were administered from October through December, and posttests from May through August; the average time between tests was approximately 8 months.

McCarthy Scales of Children's Abilities. The McCarthy scales yield a general measure of intellectual functioning (the General Cognitive Index [GCI]), as well as more specific measures of verbal, perceptual performance, short-term memory, quantitative abilities, and motor coordination. It covers the developmental range from 2½ to 8½ years. The test is well standardized and psychometrically sound, with split-half and test–retest reliabilities approximating .9 for the GCI (Sattler, 1988).

PPVT-R. The PPVT-R was utilized as a major measure of language development, because of its excellent standardization, reliability (split-half and test–retest coefficients of approximately .7–.8 in the preschool age range), and validity (Sattler, 1988). Although the test measures only receptive vocabulary, using a picture-pointing format, it is substantially correlated with a variety of other language measures. Further, some form of vocabulary test is included in nearly all intelligence tests, permitting an evaluation of the prediction based on early vocabulary skills.

Participants: Follow-Up Sample

After the students graduated from the intervention phase of the study, we followed up with them at 1-year intervals and gave them a battery of tests that varied with the age of the student. For the present study, we examined the students' test scores from age 13, approximately 8 years after the children entered the intervention. This age was selected because it was the age at which the largest number of students had been tested. A total of 171 students received the battery at age 13. This group was composed of 86 students in the ML curriculum group and 85 in the DI curriculum group. Of the remaining children, some had not yet reached 13, a few reached the age of 13 before funding was available for follow-up testing, or their families could not be reached or did not agree to participate.

Because children left the intervention at different ages, the number of years that had passed since beginning the intervention varied for students at a given age during the follow-up period; however, the two groups were comparable on this measure. Table 1 includes these and other descriptive measures for the sample. A series of *t* tests confirmed no significant differences on any of these measures.

Follow-Up Measures

Peabody Individual Achievement Test (PIAT). The PIAT (Dunn & Markwardt, 1970) was selected as the primary academic measure because of its good standardization, design as an individually administered achievement test, wide range of subtests and age of applicability, and suitability to measure changes in general cognitive processes that DI is designed to facilitate (Sattler, 1988). Median test–retest reliabilities of .89 for the total score have been reported (Sattler, 1988), along with substantial correlations with other achievement tests. Although a revision of the PIAT is now available, the annual measures of academic achievement used in this study mandated continuation of use of the original PIAT.

Stanford-Binet Intelligence Scale–Fourth Edition (SBIS-IV). The SBIS-IV (Thorndike, Hagen, & Sattler, 1986) was selected as the primary cognitive measure because of its excellent standardization, good reliability and validity, wide range of subtests, applicability over a broad age range, and suitability to measure changes in general cognitive processes that ML is designed to facilitate (Sattler, 1988). Both internal consistency and test–retest reliability coefficients are above .9.

TABLE 1. Composition of Follow-Up Sample at Age 13 by Program

Description	Direct instruction[a]			Mediated learning[b]			Overall[c]		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Sex									
Boys	59			58			117		
Girls	26			28			54		
Ethnicity									
European American	42			56			98		
African American	36			21			57		
Other	7			9			16		
Age at pretest		4.82	.82		4.75	.91		4.79	.86
Entering McCarthy GCI		78.00	17.40		76.70	16.20		77.40	16.80
Years in intervention		1.69	.72		1.72	.81		1.71	.76

Note. McCarthy GCI = *McCarthy Scales of Children's Abilities* General Cognitive Index Scores (McCarthy, 1972).
[a]$n = 85$. [b]$n = 86$. [c]$n = 171$.

The SBIS-IV was administered approximately every 3 years; the results from age 12 were used in the present study.

Identification of Participants with "Errors of Prediction"

As shown in Table 2, both entering McCarthy GCI scores and PPVT-R standard scores predicted age 13 and earlier PIAT total scores, as well as age 12 SBIS-IV scores. A multiple-regression analysis confirmed that both scores contributed unique variance to the prediction. The multiple correlation for the combined prediction was .70, representing 49% of the total variance.

Next, a single multiple-regression analysis was performed, using both preschool-phase measures (McCarthy GCI and PPVT-R standard score) to predict the PIAT total score for the entire sample for whom all relevant measures were available. As part of the analysis, standardized residuals—the difference between the predicted and actual PIAT total scores, (standardized to $M = 0, SD = 1$ for the sample)—were computed. Positive residuals indicated relatively better performance than predicted by the overall regression formula, whereas negative residuals indicated relatively poorer performance than predicted. Standardized residuals of +/– 1.5 were interpreted as errors of prediction. All participants (n = 14) for whom the standardized residual exceeded 1.5 were identified as improvers, whereas those whose standardized residual fell below –1.5 (n = 9) were identified as decreasers. A third set of 23 participants was identified as a comparison group. Each participant in the improver and decreaser groups was matched to a comparison participant on the basis of GCI scores (within 2 points) from the end of the 1st year and an error of prediction of less than 1 standard deviation. Within these constraints, participants were matched as closely as possible for entering GCI. As shown in Table 3, the three groups were quite comparable with respect to preschool-phase GCI scores but diverged substantially in age 13 PIAT total scores. Follow-up Tukey HSD tests revealed that the improver group scored significantly higher than the comparison group, which in turn scored significantly higher than the decreaser group. The groups did not diverge on SBIS-IV scores, however.

Child, Program, and Family Measures

A number of measures were evaluated as predictors of outcome category: gender, ethnicity, preschool program (Direct Instruction vs. Mediated Learning), age at entry to preschool program, number of years in the program, income (based on application for free or reduced-cost lunch during the preschool period), and initial referral categories for special education (six nonexclusive categories: language, cognition, gross motor development, fine motor development, social behavior, and medical concerns). In addition to these measures that had been previously collected, three additional measures—temperament, family stability, and parental advocacy—were obtained. These measures were identified on the basis of previous research on resilience, particularly from Werner and Smith (1982), and conversations with project staff members about what they remembered about participants across this considerable time span. To minimize the possibility of bias from staff members' memories of actual test performance, the following two-part procedure was used. Two long-term members of the research project staff with substantial experience in testing and interviewing the children and interviewing the parents were given the list of 46 names of participants in the improver, decreaser, and comparison groups but were not told their group status. They were asked to write as much information about the child and his or her family as possible. Issues of temperament, family stability, and parental advocacy were specifically mentioned, but the staff members were encouraged to provide any

TABLE 2. Prediction (Pearson Correlations) of Ages 10 Through 13 PIAT Total Scores and Age 12 SBIS-IV

Predictor	PIAT total score[a]				SBIS-IV
	10 yrs	11 yrs	12 yrs	13 yrs	12 yrs
McCarthy GCI	.56*	.56*	.54*	.57*	.60*
PPVT-R standard score	.42*	.45*	.36*	.39*	.42*

Note. PIAT = *Peabody Individual Achievement Test* (Dunn & Markwardt, 1970); SBIS-IV = *Stanford-Binet Intelligence Scales–Fourth Edition* (Thorndyke, Hagen, & Sattler, 1986); McCarthy GCI = *McCarthy Scales of Children's Abilities* General Cognitive Index scores (McCarthy, 1972); PPVT-R = *Peabody Picture Vocabulary Test–Revised* (Dunn & Dunn, 1981).
[a]*n*s for these correlations range from 134 to 171.
*$p < .001$.

TABLE 3. Comparison of Outcome Groups on Intervention Phase McCarthy GCI, Age 13 PIAT Total Scores, and Age 12 Stanford-Binet Test Composite

Measure	Group						ANOVA
	Improver		Comparison		Decreaser		
	M	*SD*	*M*	*SD*	*M*	*SD*	
McCarthy GCI[a] (entering)	80.2	14.2	80.7	12.9	78.6	10.7	$F(2, 43) = .09$, *ns*
Age 13 PIAT total[a]	108.1	7.0	84.8	10.9	68.7	4.7	$F(2, 43) = 57.6$, $p < .001$
Age 12 SBIS-IV composite[a,b]	74.6	18.1	76.4	18.1	82.1	8.5	$F(2, 35) = .44$, *ns*

Note. McCarthy GCI = *McCarthy Scales of Children's Abilities* General Cognitive Index scores (McCarthy, 1972); PIAT total = *Peabody Individual Achievement Test* total score (Dunn & Markwardt, 1970); SBIS-IV composite = *Stanford-Binet Intelligence Scale–Fourth Edition* composite score (Thorndike, Hagen, & Sattler, 1986).
[a]All scores are standard scores, $M = 100$, $SD = 15$. [b]SBIS-IV scores were not available for all participants.

other information about the child and family they could remember, in case other factors emerged as relevant (though none with enough frequency to be analyzed did emerge). The narratives were then independently coded by individuals with training in child development who were unfamiliar with either the families or the original testers. After some preliminary investigations, the coding scales proved reasonably reliable. (The coding system is detailed in the Appendix.)

School Placement

Approximately 75% of the students in this project attended a single urban school district for at least part of their education. With the consent of parents and/or adult students, we obtained special education history information from the school district. Because of moves in and out of district and because the relevant records stretched back to the late 1980s, with several classification system changes having occurred since then, the information was less complete than we had hoped for. We had hoped to be able to identify students who had been placed in general education, minimal special education (consultation or research room), and self-contained special education. However, due to ambiguities in the records, we were limited to differentiating between general and special education placement for each academic year.

Results

Longitudinal Trends in Mean PIAT Performance

For the present report, PIAT total scores at age 13 were used to identify groups. It is of interest to examine longitudinal trends in PIAT scores to determine if the scores at age 13 adequately reflect overall performance and to examine the nature of developmental change. Figure 1 includes the preschool-phase measures, as well as PIAT total scores at ages 8 through 13 for the improver, decreaser, and comparison groups. As the figure shows, the scores at age 13 reflect stable developmental trends. (Note that the scores appear relatively high in absolute terms, reflecting the use of old norms for the original PIAT; a comparison of the two tests for seventh graders [Dunn & Dunn, 1981] revealed a 6.5-point mean difference between the two tests.)

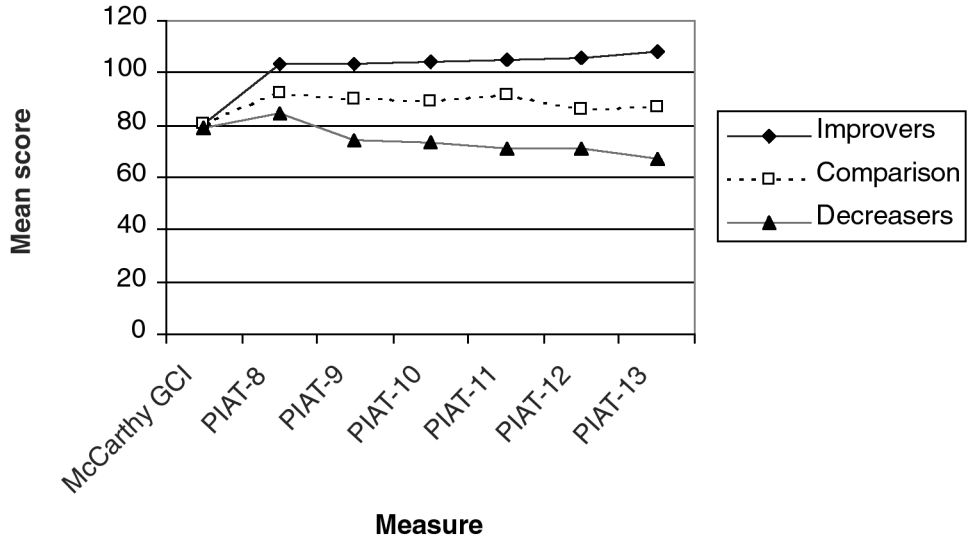

FIGURE 1. Longitudinal trends in mean PIAT performance for three outcome groups. *Note.* McCarthy GCI = *McCarthy Scales of Children's Abilities* General Cognitive Index scores (McCarthy, 1972); PIAT = *Peabody Individual Achievement Test* (Dunn & Markwardt, 1970).

Comparison of Prediction Groups

Child Variables. A comparison of the three groups on gender, ethnicity, and preschool program variables is provided in Table 4. Although there was a trend for males to have more positive outcomes, this was nonsignificant. In contrast, ethnicity and temperament were significantly related to outcome. A higher proportion of European American participants were in the improver group, and the converse was true for the African American and "other" groups. With respect to temperament, children with a relatively positive temperament (score = 1) were more likely to be in the improver group, whereas those viewed as difficult, shy, or unremarkable (score = 2 or 5) were more likely to be in the decreaser group.

Table 5 presents a comparison of the groups on the basis of initial referral category. Although there are a number of interesting trends, particularly the findings concerning children referred for social reasons, the only significant difference is that children referred for gross motor concerns were more likely to be in the improver group.

Preschool Program Variables. A comparison of the three prediction groups on preschool program variables is also included in Table 4. There was no relation between outcome and preschool program (DI or ML), age of entry, or the number of years the child was in the preschool program.

Family Measures. Table 6 summarizes comparisons of the three groups on income, family stability, and parental advocacy. There is a nonsignificant trend for lower income, as indexed by application for free or reduced-cost lunch, to be associated with outcome as a decreaser. Neither family stability nor parental advocacy appeared to be related to outcome.

Risk/Protective Factors Model. On the basis of the above analyses, three variables appeared to be significantly related to outcome group: temperament, ethnicity, and referral for gross motor concerns. For three others there was a suggestive nonsignificant trend: sex, income, and (absence of) referral for social concerns. Because a case can be made for the importance of these variables on the basis of theory and previous research, we decided to include them in the model. A summary score of "protective factors" was computed by adding one point for each of the following traits: generally positive temperament, male, European American ethnicity, middle income, referral for gross motor concerns, and absence of referral for social concerns. The summary score is significantly associated (Spearman $r = .50$, $p < .001$) with outcome group treated as an ordered, three-valued variable (where 1 = decreaser and 3 = improver). Figure 2 illustrates the probability of placement in the three outcome groups as a function of the summary score.

School Placement

At ages 8 through 13, school placement information was available for between 24 and 28 of the students. Figure 3 illustrates the proportion of students (for whom information was available) in the three outcome groups who had special education placement at each age. Students in the improver group were consistently less likely to be in special education than were students in the decreaser and comparison groups,

TABLE 4. Comparison of Outcome Groups on Child and Preschool Program Variables

	Group									
	Improver			Comparison			Decreaser			
Measure	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*p*[a]
Sex										*ns*
Boys	12			16			8			
Girls	2			7			3			
Ethnicity										< .05
European American	10			14			2			
African American	4			7			3			
Other	0			2			4			
Temperament category										< .05
1	7			11			0			
2	5			6			6			
4	0			0			0			
5	0			5			2			
6	2			0			1			
Program										*ns*
DI	5			10			4			
ML	9			13			5			
Age at entry		4.93	1.29		5.09	.92		4.76	.74	*ns*
Yrs. in program		1.93	.73		1.65	.83		1.89	.78	*ns*

Note. DI = direct instruction; ML = mediated learning. Temperament category: 1 = positive, 2 = not remarkable, 4 = hostile or aggressive, 5 = otherwise difficult, 6 = diagnosed disorder or syndrome (see Appendix).
[a]Chi-square test for categorical data; one-way analysis of variance for interval data.

TABLE 5. Initial Referral Categories (Nonexclusive) as Predictors of Outcome Category

		Group						
		Improver		Comparison		Decreaser		
Referral category	*n*	*n*	(%)	*n*	(%)	*n*	(%)	*p*[a]
Language								
Referred	34	12	(35)	15	(44)	7	(21)	*ns*
Nonreferred	12	2	(17)	8	(67)	2	(17)	
Cognition								
Referred	11	4	(36)	4	(36)	3	(27)	*ns*
Nonreferred	35	10	(29)	19	(54)	6	(17)	
Social								
Referred	25	7	(28)	11	(44)	7	(28)	$\chi^2(2) = 2.49$
Nonreferred	21	7	(33)	12	(57)	2	(10)	$p < .15$
Gross motor								
Referred	27	11	(41)	14	(52)	2	(7)	$\chi^2(2) = 7.27$
Nonreferred	19	3	(16)	9	(47)	7	(37)	$p < .05$
Fine motor								
Referred	27	11	(41)	12	(44)	4	(15)	$\chi^2(2) = 3.44$
Nonreferred	19	3	(16)	11	(58)	5	(26)	$p < .09$
Medical								
Referred	8	4	(50)	3	(25)	1	(9)	*ns*
Nonreferred	38	10	(26)	20	(53)	8	(21)	

[a]Chi-square test of significance of association between each referral category (referral vs. nonreferral) and placement in improver, comparison, or decreaser group.

TABLE 6. Comparison of Outcome Groups on Family Variables

	Group						
	Improver		Comparison		Decreaser		
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*
Income[a]	2.07	(1.00)	2.48	(.93)	1.56	(.88)	$F(2, 41) = 3.1$, $p < .06$
Family stability[b]	1.64	(.74)	1.77	(.69)	2.00	(.87)	$F(2, 42) = .64$, *ns*
Parental advocacy[b]	1.64	(.50)	1.73	(.46)	1.78	(.44)	$F(2, 42) = .26$, *ns*

[a]Three-point scale, with higher values indicating more favorable values. [b]Three-point scale, with lower values indicating more favorable values.

which did not appear to differ substantially between themselves. A composite measure of special education experience was then computed as the proportion of years for which information was available in which the student had been placed in special education. This measure could be computed for 30 of the 46 students (because missing information at any one age did not remove the student overall) and had values of 0 through 1.0, with a median value of .50. The three groups differed in the expected direction: for improvers, $M = .30$ ($SD = .41$); for comparison, $M = .62$ ($SD = .43$); and for decreasers, $M = .73$ ($SD = .30$). A univariate analysis of variance yielded a nearly significant effect for group, $F(2, 27) = 2.70$, $p < .09$. To attempt to control for individual differences in early ability measures and to provide a more rigorous test of the predictive value of special education placement beyond the positive factors identified above, a multiple regression was conducted, with PIAT total scale score at age 13 as the dependent variable. The student's McCarthy GCI score on entering the preschool program and his or her McCarthy GCI score at the end of the first year of preschool were both entered as the first block; the composite "positive factors" score defined above, which captured early child and family characteristics, was entered second; and the composite measure of special education experience was entered as the third and final block. The special education measure produced an R^2 change of .204, $F(1, 25) = 7.90$, $p < .01$.Thus, even taking into account early measures of ability and response to preschool program, along with selected child and family characteristics already demonstrated to be associated with outcome, special education is significantly and negatively related to academic achievement at age 13 (standardized beta coefficient = –.473).

Discussion

Identification of Outcome Groups

The first noteworthy issue is the meaningfulness of categorizing participants as improvers or decreasers. The criterion used for this purpose, 1.5 standard deviations for the residual of prediction, is based entirely on variance within the present sample. It therefore cannot be interpreted as having any absolute significance. At best, we can say that these participants were further from the predicted scores than the remainder of the sample. However, the finding that this classification is reliably associated with specific child and family characteristics does provide a kind of convergent validity. Furthermore, the high stability of PIAT scores from year to year (see Table 3) suggests that the classification has substantial test–retest reliability.

Strikingly, the three groups did not differ in their age 12 SBIS-IV composite scores. This difference in outcome classification for the academic and cognitive measures cannot be attributed to differential reliability; PIAT and SBIS-IV scores are predicted about equally well by the preschool tests (see Table 2). Instead, the difference between the outcome groups is genuinely specific to academic achievement. Thus, the error of prediction is not just from early test scores to later academic achievement; it also occurs between concurrent cognitive and academic measures. The divergence between the two suggests that there are powerful aspects of experience that affect academic achievement relatively independent of IQ scores.

Improvers, or Less Valid Early Testing?

The analyses and interpretation offered in the previous section are based on the assumption that errors of prediction are primarily a reflection of changes in achievement of the participants over time. An alternative interpretation is that improvers' early test scores underestimated their cognitive functioning, which provided a greater scope for improvement. Conversely, decreasers' early test scores may have overestimated their functioning.

This alternative interpretation is unlikely to be responsible for the main body of results. Improvers were more likely to be middle-class European American children of positive temperament who were not initially referred for social reasons. These four factors should have made them easier for the (middle class, European American) testers to assess accurately. Improvers were also more likely to be boys and to have been referred for gross motor concerns; these two factors could have affected testing accuracy, but they were likely to

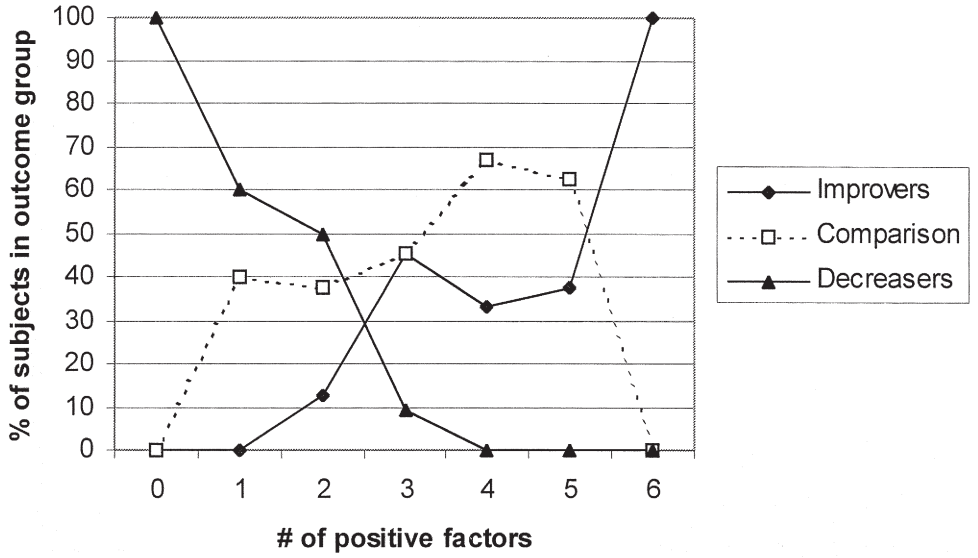

FIGURE 2. Number of positive factors and outcome status.

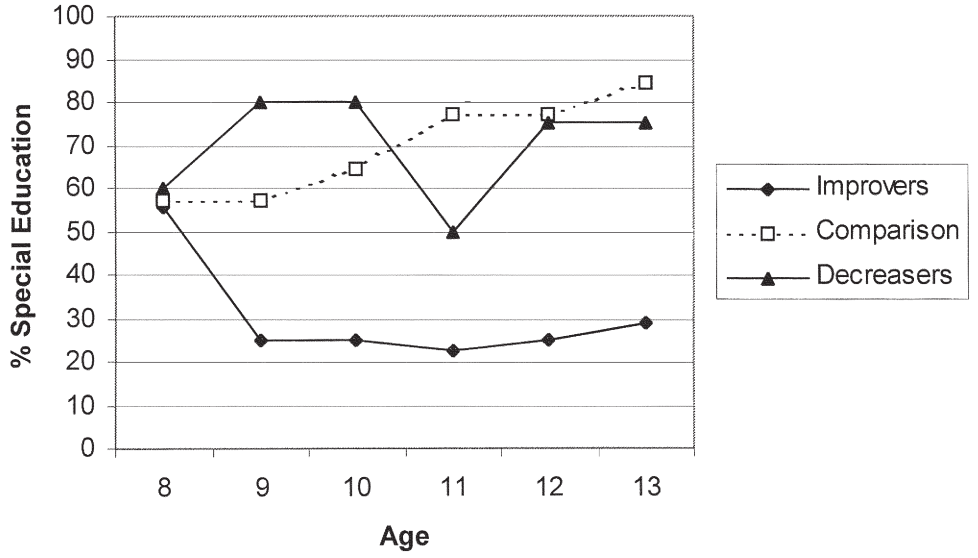

FIGURE 3. School placement experience for three outcome groups.

be outweighed by the four previous factors. We conclude that our classification to some substantial extent reflected developmental trajectory.

Factors Influencing Developmental Trajectory

The present study was originally designed as a program comparison focused on cognitive and academic measures, not as a long-term developmental study. In consequence, we have only a relatively impoverished set of child and family variables; this is especially true for social and personality factors. A measure of maternal education would be better than qualification for free/reduced-cost lunch as a measure of socioeconomic status, for example. For this reason, these results should be viewed as an underestimate, a lower bound, for the number and role of relevant child and family characteristics. Another limitation to the conclusions is that the first assessment of the children was at approximately age 4. By this time, some of these factors may have already influenced development, reducing the potential error of prediction, that is, again leading to an underestimate of effects.

Despite these limitations, the findings are intriguing in several respects. First, there are some identifiable factors re-

lated to outcome status, and they are in general plausible ones. Some of the specific findings might have been expected, even if they are discouraging reminders of the impact of sociocultural factors (e.g., the role of income/social class and European American ethnicity). The findings for temperament are notably strong and consistent with other research on resilience: None of the decreasers were viewed as likeable or outgoing, but half of the increasers were. Initial nonreferral for social concerns likely functions in the same way as temperament. We did not anticipate the relative advantage for male gender, however; in fact, this finding was in opposition to the results for gender found in the studies reviewed by Rutter (1990), all of which were conducted with at-risk children. It is frequently noted that the disproportionate representation of boys in samples with language and learning disabilities is much greater in clinic-referred samples than in community samples. A common explanation for this is that girls who are referred for special education services have to have more severe, or broader, disabilities in order to be referred. To the extent that this is correct, early measures for girls may underestimate the degree of their problem and thus fail to signal the lower probability of improving later. Further, given the results of Rutter et al. (1975) and Werner and Smith (1982), it is surprising that neither family stability nor parental advocacy appeared to play a role in outcome. It may be that because of this was conducted in an urban, relatively well-educated setting where there was a network of support agencies and institutions, all the families had sufficient social support to avoid negative effects. Taken together, these findings confirm the expectation that although there are some common factors such as positive temperament, different constellations of protective factors will be found for children with early developmental impairments. The present sample is being followed until age 18, and it is possible that the nature of the predictors will change by that time.

Second, no one factor is predominant; instead, it appears to be the conjunction of several factors that make it possible to "beat the odds," as Figure 2 illustrates. All but one of the improvers had three or more positive factors, and all but one of the decreasers had two or fewer. No one factor comes even nearly close to discriminating these two groups as well as the whole set does. It is most likely that what all these factors have in common is that they influence treatment by parents, teachers, peers, and others over the years. The combined role of nonreferral for social concerns and positive temperament is especially interesting in the light of the recent longitudinal findings of Caprara, Barbaranelli, Pastorelli, Bandura, and Zimbardo (2000). On the basis of structural equations modeling, they concluded that the strongest third-grade predictor of both academic achievement and peer social acceptance in the eighth grade was prosocial behavior, based on a combination of self-, teacher, and peer ratings. The strength of this prediction held up even when third-grade academic achievement was controlled statistically. Caprara et al. concluded that a prosocial orientation fosters mutually supportive social and intellectual relationships with peers and reduces vulnerability to depression and aggression, both of which can undermine academic motivation.

The fact that the divergence among the three outcome groups was specific to academic achievement highlights the specific role of educational experience. This discrepancy between cognitive and academic outcomes has been noted in other situations, despite the fact that intelligence tests were originally developed with the primary goal of predicting academic achievement. For example, Hart and Risley (1995) assessed 29 of the 42 typically developing children who had been in their longitudinal study as toddlers when they were 9 to 10 years of age. They found that age 3 SBIS-IV scores were substantially correlated with later standardized tests of language development but were not correlated with academic achievement in reading, writing, spelling, and arithmetic. We suggest that part of the explanation will center on the experiences children have at school and the way those experiences are affected by characteristics of the children.

The most problematic of our results concerns school placement experience. Lower academic achievement, and classification as a decreaser, was associated with greater experience in special education than was classification as an improver, even when controlling for preschool period and family measures. Special education placement is potentially both an outcome measure and a causal factor in academic achievement, and for this reason the interpretation of the association is extremely difficult. The present results are entirely consistent with the hypothesis that poor academic achievement in the early years, beyond that predictable by preschool period measures, and especially when accompanied by difficult temperament and behaviors, leads to special education placement as well as to poor academic achievement at age 13. They are equally consistent with the hypothesis that extended time in special education limits the growth of academic achievement. It is again notable that the groups did not differ in their age 12 Stanford-Binet scores. This set of results suggests that both student motivation and teacher expectation merit further examination.

Conclusions

As acknowledged earlier, the present study is inherently retrospective. There is a crucial need for long-term, prospective outcome studies of children who have been identified as having special needs in early childhood. These studies should set academic and cognitive development in a broad social, familial, and cultural perspective, in order to identify more comprehensively and more accurately the environmental factors that can have a powerful effect on the path of development. We believe that the present results do provide strong, if indirect, evidence that experience can make a powerful difference. However, characteristics of the child and family may have quite

unintentional, deleterious effects on the child's environment. The goal of education and other interventions must be to ensure that every child receives the best possible experience, to take the upwardly divergent path. Brooks has highlighted the role of "islands of competence" in social, mechanical, athletic, and other domains in building self-esteem, hope, and self-discipline, with a number of practical suggestions. Rutter (1990), drawing on theories of development as well as empirical research, has offered a particularly valuable summary:

> The limited evidence available so far suggests that protective processes include (a) those that reduce the risk impact by virtue of effects on the riskiness itself or through alteration of exposure to or involvement in the risk, (b) those that reduce the likelihood of negative chain reactions stemming from the risk encounter, (c) those that promote self-esteem and self-efficacy through the availability of secure and supportive personal relationships or success in task accomplishment, and (d) those that open up opportunities. . . . Particular attention needs to be paid to the mechanisms operating at key turning points in people's lives, when a risk trajectory may be redirected onto a more adaptive path. (pp. 209–210)

AUTHORS' NOTE

The Longitudinal Comparison Project is an ongoing study conducted at the Department of Special Education, University of Washington. It is supported by a grant from the U.S. Department of Education (H324C9990012; "A Longitudinal Follow-Up of Graduates From Two Contrasting Instructional Models: Phase 3").

REFERENCES

Brooks, R. B. (1999). Fostering resilience in exceptional children: The search for islands of competence. In V. L. Schwean & D. H. Saklofske (Eds.), *Handbook of psychosocial characteristics of exceptional children* (pp. 563–585). New York: Kluwer.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11,* 302–306.

Cole, K. N., Dale, P. S., Mills, P. E., & Jenkins, J. R. (1993). Interaction between early intervention curricula and student characteristics. *Exceptional Children, 60,* 17–28.

Dale, P. S., & Cole, K. N. (1988). Comparison of academic and cognitive programs for young handicapped children. *Exceptional Children, 54,* 439–447.

Dunn, L., & Dunn, L. (1981). *Peabody picture vocabulary test–Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L., & Markwardt, F. (1970). *Peabody individual achievement test*. Circle Pines, MN: American Guidance Service.

Englemann, S. (1980). *Direct instruction*. Englewood Cliffs, NJ: Educational Technology Publications.

Englemann, S., & Bruner, E. C. (1974). *Distar reading I and II*. Chicago: Science Research Associates.

Englemann, S., & Carnine, D. (1975). *Distar arithmetic I and II*. Chicago: Science Research Associates.

Englemann, S., & Osborne, J. (1976). *Distar language I and II*. Chicago: Science Research Associates.

Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability.* Baltimore: University Park Press.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.

Lazar, I., & Darlington, R. (1982). Lasting effects of early education: A report from the consortium for longitudinal studies. *Monographs of the Society for Research in Child Development, 47*(2–3, Serial No. 195).

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist, 53,* 205–220.

McCall, R. B., Appelbaum, M. I., & Hogarty, P. S. (1973). Developmental changes in mental performance. *Monographs of the Society for Research in Child Development, 38*(Serial No. 150).

McCarthy, D. (1972). *McCarthy scales of children's abilities*. San Antonio, TX: Psychological Corp.

Mills, P. E., Dale, P. S., Cole, K. N., & Jenkins, J. R. (1995). Follow-up of children from academic and cognitive preschool curricula at age 9. *Exceptional Children, 61,* 378–393.

Notari-Syverson, A. R., Cole, K. N., Osborne, J., & Sherwood, D. (1996). What is this? What did we just do? How did you do that? Teaching cognitive and social strategies to young children with disabilities in integrated settings. *Teaching Exceptional Children, 28,* 12–16.

Osborne, J., & Sherwood, D. (1984). *Mediated learning program for young children*. Unpublished curriculum, University of Washington, Seattle.

Rutter, M. (1990). Psychosocial resilience and protective mechanisms. In J. Rolf, A. S. Masten, D. Cicchetti, K. H. Neuchterlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology* (pp. 181–214). New York: Cambridge University Press.

Rutter, M., Yule, B., Quinton, D., Rowland, O., Yule, W., & Berger, M. (1975). Attainment and adjustment in two geographical areas: III. Some factors accounting for area differences. *British Journal of Psychiatry, 126,* 520–533.

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

Thorndike, R., Hagen, E., & Sattler, J. (1986). *The Stanford-Binet intelligence scale: Fourth edition*. Chicago: Riverside.

Werner, E. E. (1999). Risk and protective factors in the lives of children with high-incidence disabilities. In R. Gallimore, L. P. Bernheimer, D. L. Macmillan, D. L. Speece, & S. Vaughn (Eds.), *Developmental perspectives on children with high incidence disabilitie*s (pp. 15–31). Mahwah, NJ: Erlbaum.

Werner, E., & Smith, R. S. (1982). *Vulnerable but invincible: A longitudinal study of resilient children and youth*. New York: McGraw-Hill.

Appendix

The narratives produced by the two primary research project staff members for each participant were coded independently by other individuals, who were unaware of the outcome group to which the participant had been assigned. Three aspects of child and family characteristics were coded.

Temperament

1 = outgoing, likeable, relates well, easygoing, or other positive
2 = not remarkable, no information, or shy/withdrawn
3 = this code was not used
4 = hostile or aggressive
5 = not hostile or aggressive but otherwise difficult, hyperactive, or impulsive
6 = diagnosed mental disorder or congenital syndrome

Agreement = 73%, Cohen's kappa = .62

Family Stability

1 = high stable family, could be single parent or early adoption, if stable afterward
2 = low/moderate changes, such as single divorce/remarriage, or little or no information available
3 = several moves or partner changes, includes foster care placement, running away from home

Agreement = 73%, Cohen's kappa = .59

Parental Advocacy

1 = unusually strong, such as getting the child a special tutor/counselor or home schooling
2 = not remarkable or no information
3 = appears low

Agreement = 89%, Cohen's kappa = .76