Age versus Schooling Effects on Intelligence Development

Sorel Cahan and Nora Cohen
The Hebrew University of Jerusalem

CAHAN, SOREL, and COHEN, NORA. *Age versus Schooling Effects on Intelligence Development.* CHILD DEVELOPMENT, 1989, **60**, 1239–1249. The effect of formal education, as opposed to chronological age, on intelligence development has suffered from inadequate empirical investigation. Most studies of this issue have relied on natural variation in exposure to school among children of the same age, thus confounding differences in schooling with differences in other intelligence-related variables. This difficulty can be overcome by a quasi-experimental paradigm involving comparison between children who differ in both chronological age and schooling. The present study applies this paradigm to the estimation of the independent effects of age and schooling in grades 5 and 6 on raw scores obtained on a variety of general ability tests. The sample included all students in Jerusalem's Hebrew-language, state-controlled elementary schools. The results unambiguously point to schooling as the major factor underlying the increase of intelligence test scores as a function of age and to the larger effect schooling has on verbal than nonverbal tests. These results contribute to our understanding of the causal model underlying intelligence development and call for reconsideration of the conceptual basis underlying the definition of deviation-IQ scores. Some implications of these results concerning the distinction between intelligence and scholastic achievement, the causal model underlying the development of "crystallized" and "fluid" abilities, and the notion of "culture-fair" tests are discussed.

From the beginning of intelligence testing, the concept of intelligence has been considered a developmental one, closely related to chronological age (Binet & Simon, 1916; Reynolds, 1982; Sternberg & Powell, 1983; Wohlwill, 1980). As pointed out by Jensen (1980), the very construct of general intelligence implies a systematic growth in mental ability from infancy to maturity. Thus, raw scores on tests of general intelligence should and do show a positive regression on chronological age during childhood (e.g., Wechsler, 1974).

However, while intelligence theory is quite explicit concerning the developmental nature of mental ability, the underlying causal model is much less clear. Obviously, this is due to the inextricable covariation between the factors affecting cognitive development (e.g., biological maturation, accumulation of experience) and chronological age. Indeed, in normal developmental conditions, chronological age actually stands for both biological and psycho-educational development and provides the only available scale for the measurement of informal learning and accumulation of experience. Therefore, the distinction between chronological age and most age-related variables is both analytically problematic and empirically possible only under extreme physical or psychological environmental conditions. Clearly, experimenting with such conditions is ethically unacceptable.

Schooling (last grade completed) is a notable exception in this respect. Unlike most other age-related factors, it is not inherent to chronological age. Moreover, an analytical distinction between schooling and other factors that may affect intelligence development is desirable on theoretical grounds: school is explicitly aimed at the development of intellectual abilities. Indeed, the learning processes involved in many school activities are thought to affect the formation of the cogni-

This research was supported by the Israel Foundations Trustees, by the Fund for Basic Research Administered by the Israel Academy of Sciences and Humanities, by the Martin and Vivian Levin Center for the Normal and Psychopathological Development of the Child and Adolescent, the Hebrew University, and by the NCJW Research Institute for Innovation in Education, School of Education, The Hebrew University. We gratefully acknowledge the assistance of Lavee Artman, Daniela Azar, Galit Ganon, Gabriela Reem, Anat Sarel, and Jacob Yellinek in data collection and analysis. We also appreciate the graphics work of Kari Druck and the editorial revisions of Helene Hogri. Finally, we thank three anonymous reviewers for comments on a draft of this article. Requests for reprints should be addressed to Sorel Cahan, School of Education, The Hebrew University, Jerusalem 91905, Israel.

[*Child Development*, 1989, **60**, 1239–1249. 0009-3920/89/6005-0023$01.00]

tive strategies needed for successful performance on general ability tests (Glaser, 1984). This has led to an increasing recognition of the potential effect of formal education on the development of intelligence and on the corresponding increase of mean raw test scores as a function of age (e.g., Anastasi, 1986; Cattell, 1963, 1971; Cronbach, 1984; Horn, 1970, 1978).

Unfortunately, however, the empirical evidence concerning this issue is inconclusive. Owing to the impossibility of experimenting with school attendance, the research approach to the investigation of schooling effects has heretofore been of a post-hoc nature, relying on natural variation in schooling among people of the same age. Since school attendance at the elementary level is compulsory in modern societies, such variation could be found, with few exceptions, only: (*a*) among adolescents and adults who reached different levels of secondary and higher education (e.g., Harnqvist, 1968); (*b*) following accidental instances of a temporary lack of schooling in specific locations (e.g., deGroot, 1951); or (*c*) in developing countries where school attendance was not universal even at the elementary level (e.g., Scribner & Cole, 1981).

However, all these studies were affected by the same major methodological problem—selection, that is, the possible confounding of differences in schooling with those in other intelligence-related variables. Madaus, Airasian, and Kellaghan (1980), reviewing such investigations, conclude: "While suggesting the importance of schooling, these studies do not control for the operation of other factors which might have affected children's performance" (p. 47). Unfortunately, statistical control (e.g., covariance analysis, multiple regression) may over- or undercorrect for correlates of schooling, leaving us in doubt as to its true effect (Cook & Campbell, 1979).

We believe that additional research that can better cope with the problem of selection is needed. In pursuing this course, we have made use of the recently proposed "between-grade level" approach (Cahan & Davis, 1987). According to the rationale underlying this approach, the overall cross-sectional increase in mean raw scores as a function of age is decomposed into within-grade and between-grade segments, which can be unambiguously attributed to age and schooling effects, respectively. This is based on two assumptions: (1) The "allocation" of children to birth dates is random. (2) Grade level is solely a function of chronological age, that is, admission to school is based on chronological age only, according to some arbitrary cut-off point, and progression through grades is automatic (i.e., there are no dropouts and children are neither kept back nor skipped).

On the basis of these assumptions, the net effects of chronological age and schooling obtain from differences in mean raw scores between extreme age groups. The difference between the mean test scores of the oldest and youngest children in each grade gives an estimate of the net effect of 1 year difference in chronological age in that grade, while that between the youngest children in any given grade level (X) and the oldest children in the lower adjacent grade level (X – 1) provides an estimate of the effect of 1 year of schooling (the 1-day age difference between these groups is negligible). If the tests are administered at the end of the school year, then the estimate refers to the effect of 1 year of schooling in grade X.

The estimation of age and schooling effects via differences between the mean test scores of the extreme birth dates in each grade has a serious shortcoming. There are budgetary and logistical constraints in locating and testing a sufficient number of subjects born on these extreme dates to reduce the expected magnitude of random differences between birthdays—both within and between grades—and, thus, the standard errors of the estimated effects. Note that the problem cannot be solved by using larger birthdate categories (e.g., months), since the difference between the mean test scores of the youngest children in the higher grade and the oldest children in the lower grade would then no longer be attributable solely to schooling.

One way to cope with this problem is to base the estimation of the independent effects of age and schooling on estimates of the population mean test scores of the youngest and oldest children in each grade, which are more reliable than empirically obtained sample means. The regression discontinuity design (Cook & Campbell, 1979) provides an appropriate method to achieve this aim. Applied to our context, this design involves linear prediction of the mean test scores of the youngest and oldest children in each grade by means of the best-fitting regression line of test scores on chronological age across the entire age range in that grade.[1] Higher precision of the estimated effects is thus achieved by use of

[1] Even though the within-grade increase in mean raw test scores as a function of age need not be linear, its approximation by linear regressions is amply satisfactory due to the short time intervals in each grade.

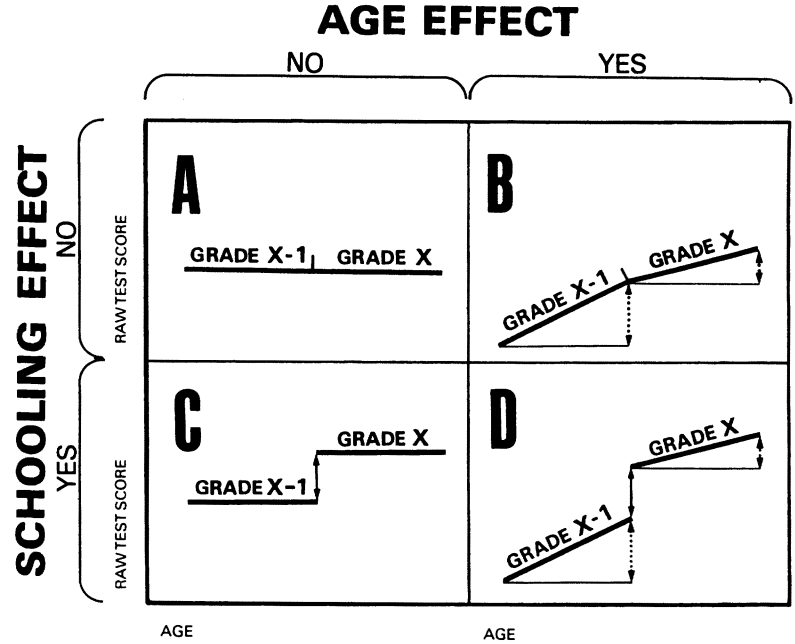

FIG. 1.—The independent effects of age (dotted arrows) and schooling (solid arrows) in the between-grades regression discontinuity design. Four hypothetical examples.

this readily available information rather than by increasing sample size in the extreme birthdays. The effect of age is reflected, in the regression discontinuity design, in the slope of the within-grade regressions of test scores on chronological age, while the effect of schooling is reflected in the discontinuity between them (see Fig. 1). Specifically, the estimated effect of a 1-year difference in chronological age in a given grade (X) equals the difference between the oldest and youngest students in that grade in mean predicted scores (see dotted arrows in Fig. 1), while the estimated effect of 1 year of schooling equals the difference in mean predicted scores between the youngest children in that grade and the oldest children in the lower adjacent grade (X − 1). This design is applied here to the estimation of the effects of schooling in grades 5 and 6 on scores obtained on widely used group tests of general ability.

Method

Measures

Tests.—Twelve tests, covering a wide range of item content (e.g., analogies, series, sentence completion, vocabulary) and varying in item modalities (verbal, numerical, figural), were selected from well-known group tests of general ability: the Cognitive Ability Test (CAT; Thorndike & Hagen, 1971), Milta—a Hebrew version of the Lorge Thorndike Test (Ortar & Shachor, 1980), Standard Progressive Matrices (Raven, 1983), and Cattell and Cattell's (1965) Culture Fair Intelligence Test (see Table 1). Tests 1 and 3 were translated into Hebrew. Tests 6–12 were taken from the source battery in the original form, including time limits, while long tests (tests 1–5) were abridged to meet a 4-min time limit. In order to avoid floor and ceiling effects, final length was determined on the basis of a pilot study. The resulting decrease in the reliabilities of these tests has no serious consequences in this study, which is concerned with differences between means (Cahan & Davis, 1987; Stanley, 1971). Twelve raw scores (number of correct answers), one for each test, were computed for each subject.

Chronological age.—Before test administration, exact birth date (day, month, and year) of each subject was obtained from school files.

Subjects

The target population of the study consisted of all the fourth, fifth, and sixth graders attending Jerusalem's Hebrew-language, state-controlled elementary schools in 1987 (with the exception of schools for special education). Sixty-one out of these 62 schools, with a total population of 12,090 fourth, fifth, and sixth graders (exclusive of in-school special education students and new immigrants), agreed to cooperate. Ninety-one percent of

TABLE 1

DESCRIPTION OF THE TWELVE TESTS (in Order of Presentation)

Test No. and Name	Source[a]	No. of Items Selected	Time Limit in the Study (min)
1. Verbal classification	CAT (54)	16[b]	4
2. Figure classification	CAT (45)	18	4
3. Verbal analogies	CAT (45)	21[b]	4
4. Figure analogies	CAT, Figure analysis (45)	20	4
5. Matrices	SPM, Sets C, D (24)	16	4
6. Vocabulary	Milta (40)	40	4
7. Number series	Milta (18)	18	5
8. Figure series	CFIT, Series (12)	12	3
9. Verbal oddities	Milta (19)	19	4
10. Figural oddities	CFIT, Classifications (14)	14	4
11. Word arithmetic problems	Milta (18)	18	5
12. Sentence completion	Milta (23)	23	4

[a] Source battery name: CAT—Cognitive Ability Test, levels A–E; SPM—Standard Progressive Matrices; CFIT—Culture Fair Intelligence Test, Scale 2, Form A; Milta—Milta, Elementary Level (grades 4–6). Number of items in parentheses.

[b] Two items constructed for this study have been added to test 1 and one item to test 3.

the students (11,099) attended school on the day of test administration. Of these, 93% took all 12 tests, 96% took at least 11 tests, and 98% at least 10 tests. The participation rate was stable across grade levels and did not vary considerably between schools.

Procedure

The tests were administered in a fixed order (see Table 1) on a classroom basis between May 8 and June 8, 1987, during a 3-hour morning session with a 30-min break. Two testers (Hebrew University students who had received special instruction) were present in each classroom. A few days prior to the administration of the tests, each class was given general explanations about the item format and the response sheet. In addition, the administration of each subtest was preceded by a short explanation and two illustrative examples of the particular task.

Data Analysis

Implementation of the between-grades model.—The truth of the first assumption underlying the between-grades paradigm—namely, the random allocation of children to birth dates—cannot be empirically tested. However, since we tested students in only three adjacent grades in a relatively homogeneous population, this assumption seems reasonable with respect to the between-grade variability. As far as the within-grade randomization is concerned, exceptions to this assumption are not likely to affect the estimation of the age and schooling effects unless they are monotonically related to birth date.

The second assumption of the model—namely, that admission to school is based solely on chronological age and that grade progression is automatic—is only partially true. While grade retention and grade skipping are seldom practiced in the Israeli elementary educational system, admission to school is sometimes delayed and sometimes accelerated. In any given grade, therefore, there are children whose age should place them in a lower or higher grade, and there are others who are "missing" (i.e., are learning in a higher or lower grade). More important, the delay or acceleration of school admission is not random. The children whose admission was delayed are likely to be less developed intellectually than the other children in their age group, and those whose admission was accelerated are likely to be more developed. In addition, the relative frequency of grade misplacement is likely to be related to month of birth, being particularly high near the cutoff point. Cook and Campbell (1979) discuss this possibility as a "fuzzy cutting point." The empirical data strongly support this prediction. Delays are especially frequent among the oldest children in each cohort and accelerations among the youngest (see Fig. 2). Note that both groups are born in December. This is because the official deadline for admission to school in the Israeli educational system is defined according to the Jewish calendar as Teveth 1, which usually falls in December, the exact Gregorian date varying from year to year. This explains the appearance of December at both extremities of the appropriate age range in each grade.

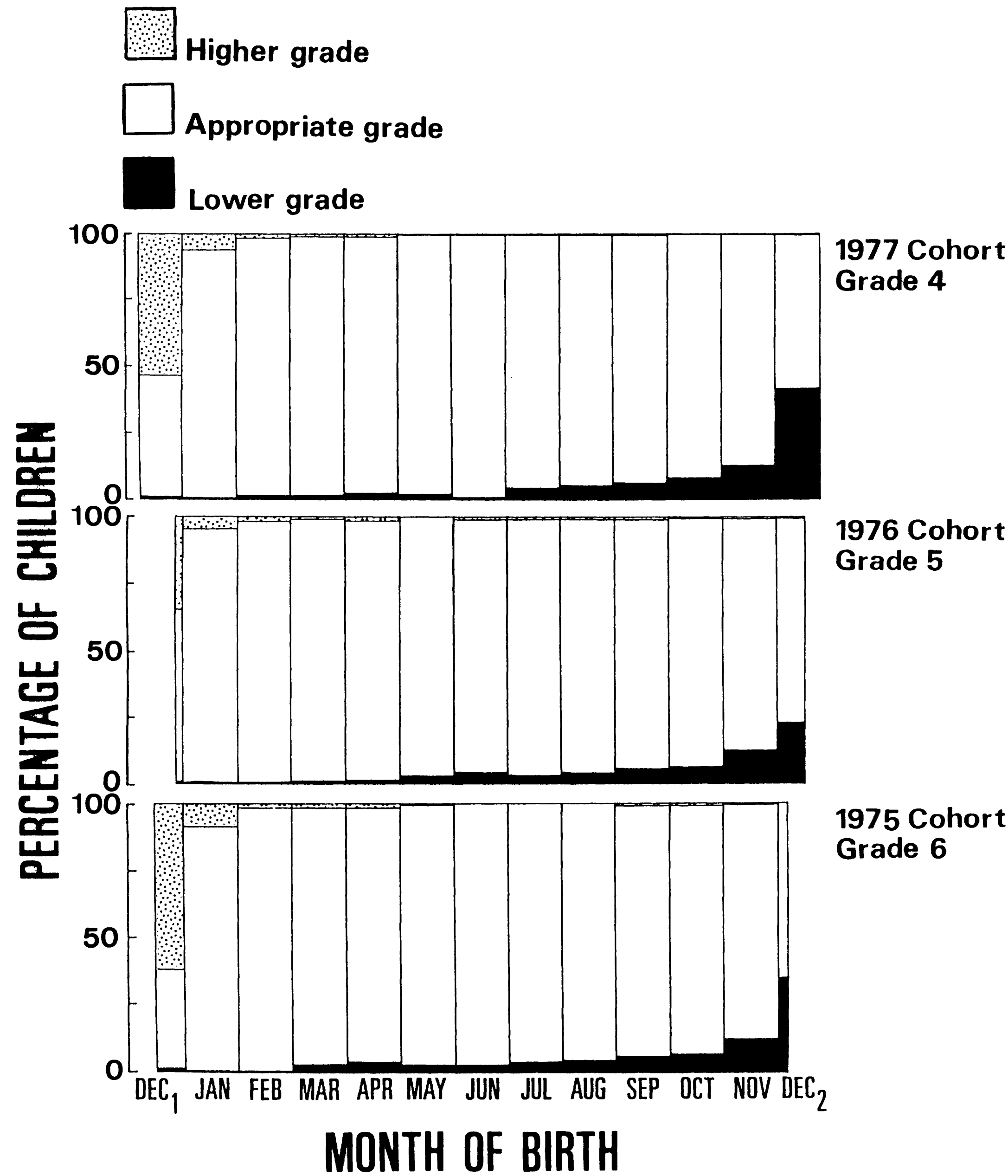

FIG. 2.—Grade placement as a function of month of birth for the three cohorts. DEC_1 denotes December of the previous year. The proportion of placements in higher grades in the 1975 cohort (grade 6) and the proportion of placements in lower grades in the 1977 cohort (grade 4) are estimates based on the other two cohorts.

There are two ways in which selective misplacement may affect the within-grade regression slopes: (1) Due to the existence of underage and overage children in each grade. The direction of this effect cannot be established a priori since age and selection counteract each other in this case: the underage children are also brighter, while the overage ones are generally duller. (2) Due to the "missing" children in each grade. At the lower extreme of the age range the missing children are those that have been delayed; hence, the mean test score of the remaining children in the youngest groups is higher than the "true" one. At the higher extreme of the age range, selection operates in the opposite

TABLE 2

RAW SCORE POOLED-WITHIN-AGE STANDARD DEVIATIONS OF THE 12 TESTS BY GRADE LEVEL

TEST NUMBER AND NAME	GRADE 4	GRADE 5	GRADE 6
1. Verbal classification	3.9	3.6	3.5
2. Figure classification	3.2	3.1	3.1
3. Verbal analogies	4.7	4.5	4.3
4. Figure analogies	4.3	4.3	4.2
5. Matrices	3.0	3.2	3.3
6. Vocabulary	5.7	6.2	6.7
7. Number series	3.2	3.1	3.1
8. Figure series	2.3	2.2	2.1
9. Verbal oddities	2.6	2.4	2.4
10. Figural oddities	1.9	1.8	1.8
11. Word arithmetic problems	2.3	2.5	2.8
12. Sentence completion	4.5	4.6	4.6

direction: the missing children are the brightest ones, whose admittance to school has been accelerated. Consequently, the mean test scores of the remaining children in the oldest group are lower than the true ones. Thus the missing children at both extremities of the age range affect the within-grade regression slopes in the same direction: the empirically obtained slope is attenuated, that is, smaller than the true one, thus leading to an underestimation of the age effect and an overestimation of the schooling effect.

In order to cope with this problem, we excluded from the computation of the within-grade regressions two groups of subjects: (*a*) students who were under- or overaged, and (*b*) students born in November or December, that is, the birth dates with the highest proportion of missing students (see Fig. 2). Thus, each within-grade regression was based only on children born between January and October of the appropriate year for that grade. While in each of these months of birth there still was a small proportion of missing students (see Fig. 2), this proportion did not vary considerably between months of birth and, therefore, was not likely to affect the within-grade slope.[2]

The effect size metric.—In order to allow for between-test comparability, the estimates of the age and schooling effects were standardized using the pooled-within-age standard deviation in grade 4. The resulting numerical values can be legitimately interpreted in terms of "effect sizes" (Glass, 1976; Hedges & Olkin, 1985). Obviously, the choice of the particular standard deviation as the common metric is arbitrary; however, choice of any other standard deviation—which may have yielded different effect sizes—would not affect the ratios between them.[3] Furthermore, in light of the small between-grade differences in the within-age variability of the 12 tests (see Table 2), this choice is inconsequential.

[2] The paradoxical nature of this solution, which excludes from the estimation of the age and schooling effects the only birth dates entering the conceptual definition of these effects, is worth noting. As a result, the actual estimation of these effects and the underlying rationale are based on entirely different age groups. Note also that this solution is applicable only in the context of the regression-discontinuity design. Only in this context can the selection-affected mean test scores of the extreme age groups be replaced by values which are "corrected for selection." Consequently, the regression discontinuity design is not only logistically more effective than the exteme-age-groups design in increasing the precision of the estimated effects of age and schooling, but, moreover, its adoption is essential for coping with the central problem of selection.

[3] Alternatively, effect sizes could be obtained by standardizing each grade-specific effect by a different standard deviation (e.g., the standard deviation of the lower grade level involved in the definitions of the effect) rather than by a common standard deviation. Conceptually, this procedure answers an entirely different question. From a practical point of view, however, the choice between these two approaches is critical only when there are considerable between-grade differences in variability. In this case, they would yield not only different absolute magnitudes of the estimated effects, but also different ratios between them.

TABLE 3

THE RELIABILITY COEFFICIENTS (Cronbach's α) OF THE 12 TESTS IN GRADE 4

Test Number and Name	Cronbach's α
1. Verbal classification..........	.87
2. Figure classification	.80
3. Verbal analogies.............	.87
4. Figure analogies.............	.86
5. Matrices	.77
6. Vocabulary..................	.86
7. Number series	.83
8. Figure series................	.71
9. Verbal oddities..............	.64
10. Figural oddities	.49
11. Word arithmetic problems....	.70
12. Sentence completion	.87

The between-grades variability of the estimated effects.—This study investigated the effects of schooling in two grade levels (fifth and sixth) and the effects of age in three grade levels on scores obtained on 12 tests. Consequently, for each test there are two grade-specific schooling effects and three grade-specific age effects to be estimated, that is, a total of 24 grade × test schooling effects and 36 grade × test age effects.

In principle, the design allows for both within-test, between-grades comparisons and within grade, between-tests comparisons in the absolute and relative magnitudes of the estimated effects of age and schooling. These comparisons, however, may yield spurious results since the variability between the empirically obtained estimates may be affected by random differences between cohorts as well as by idiosyncratic psychometric characteristics of the tests, which are grade-specific. To cope with this problem, we estimated the true effects of 1 year of age and 1 year of schooling for each test by averaging the three grade-specific age effects and the two grade-specific schooling effects, respectively.[4] This is a conservative approach, which attributes between-grade variability of the estimates to random error and is based on the generally incorrect assumption that the true effects do not vary between grades (i.e., that the within-grade regressions are parallel and equidistant). We believe, nonetheless, that its adoption is the best course of action: first, because the information that these averages may fail to reveal is relatively minor, considering that we are dealing with only three adjacent grades, and second, and more important, because the between-test comparisons are far more central to this study than the between-grade comparisons for each test.

Correction for attenuation.—Since the empirically obtained estimates of the age and schooling effects are expressed in grade 4 standard deviation units, they are attenuated by the measurement error in that grade. Therefore, the empirically obtained estimates were corrected for attenuation using the grade 4 reliability coefficients of the 12 tests (see Table 3). This correction is especially important because the extent of attenuation may vary between tests, thus affecting not only the absolute value of the estimated effect for each test but also their rank order across tests.

Results

The estimated net effects of 1 year of schooling in grades 5 and 6 and 1 year of age in grades 4–6 for each of the 12 tests are presented in Table 4. To clarify the meaning of these numerical values and their relation to the between-grades paradigm, Figure 3 gives the regression discontinuity patterns corresponding to the four most extreme cases: the lowest and highest empirically found effects of age (Fig. 3*A* and 3*B*, respectively) and schooling (Fig. 3*C* and 3*D*, respectively).

The results are unambiguous. First, they point to the general nature of the effect of schooling: schooling affects scores on all tests. Second, they point to the larger effect schooling has on verbal than nonverbal tests. Third, they point to schooling—rather than to other age-related factors—as the major factor underlying the increase of intelligence test scores as a function of age: for nine out of the 12 tests, the effect of 1 year of schooling is larger than that of 1 year of age. Moreover, in all these cases, the effect of schooling is about twice the effect of age (see col. C in Table 4).

These phenomena are best seen in Figure 4, which presents the joint distribution of the estimated age and schooling effects. Each point in the figure represents one test. The coordinates of each point are the values of the estimated net effects of 1 year of age and 1 year of schooling for the test in question. Two features of Figure 4 are worth emphasizing, namely, the clear distinction between the verbal and nonverbal tests in terms of the mag-

[4] These averages are equivalent to the coefficients of age and grade level in the across-grade multiple regression of raw test scores on age and grade level, expressed in grade 4 standard deviation units.

TABLE 4

ESTIMATED EFFECTS OF 1 YEAR OF AGE IN GRADES 4–6 AND 1 YEAR OF SCHOOLING IN GRADES 5 AND 6 ON SCORES OBTAINED ON THE 12 TESTS

	Estimated Net Effect of 1 Year of:		
Test Number and Name	Age (A)	Schooling (B)	B/A (C)
Verbal tests:			
1. Verbal classification	.12*	.23*	1.9
3. Verbal analogies	.14*	.27*	1.9
6. Vocabulary	.19*	.40*	2.1
9. Verbal oddities	.05**	.35**	7.0
11. Arithmetic problems	.16**	.50**	3.1
12. Sentence completion	.18*	.41*	2.3
Numerical tests:			
7. Number series	.15*	.26*	1.7
Figural tests:			
2. Figure classification	.16**	.16**	1.0
4. Figure analogies	.22*	.14*	.6
5. Matrices	.13**	.27**	2.1
8. Figure series	.19*	.11*	.6
10. Figural oddities	.09***	.20***	2.2

NOTE.—In pooled-within-age grade 4 standard deviation units.
* SE = .05.
** SE = .06.
*** SE = .07.

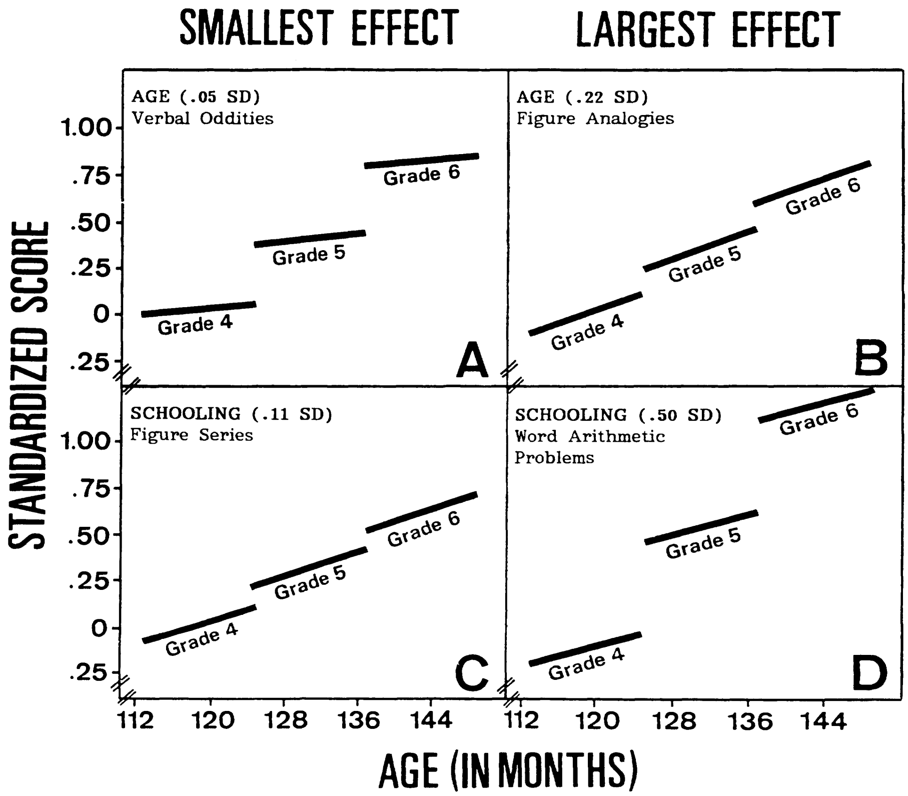

FIG. 3.—Selected examples of empirically obtained regression discontinuity patterns illustrating relatively extreme magnitudes of the age and schooling effects (in pooled-within-age grade 4 standard deviation units).

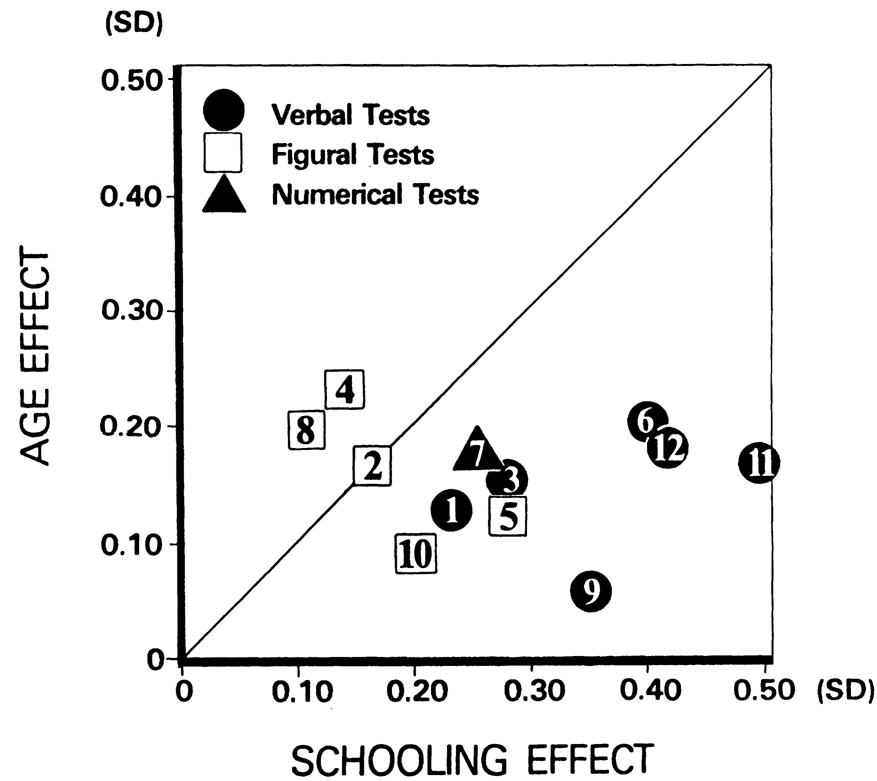

FIG. 4.—Joint distribution of age and schooling effects for the 12 tests (in pooled-within-age grade 4 standard deviation units; numbers identify tests, see Table 1).

nitude of the effect of schooling, and the fact that only two out of the 12 tests are above the main diagonal; for the remaining 10 tests, the effect of 1 year of schooling equals or exceeds the effect of 1 year of age.

Discussion

The present study has shown that schooling has a considerable effect on intelligence test scores. Of course, the magnitude of this effect may vary between educational systems, where schooling at a particular grade level may mean different things, as well as between grade levels in the same educational system. Nevertheless, we contend that the main conclusion reached here is generalizable to most educational systems. Hence we believe that our results have general implications, some of which we now elaborate.

First, they foster our understanding of the development of intelligence and point to the critical contribution of formal education to this process, thereby providing more valid empirical evidence in support of the "schooling effect" hypothesis. This contribution is particularly relevant to the explanation of IQ differences both within and between groups (e.g., racial, social, or cultural). A related implication concerns the measurement of intelligence. Historically, all the basic concepts in intelligence measurement (mental age, intelligence quotient, and the modern concept of deviation-IQ) have been defined solely on the basis of age, irrespective of schooling. From this, it may be inferred that test authors and users assume the independent effect of schooling on intelligence test scores to be negligible (Angoff, 1984; Flanagan, 1951). The evidence provided by this study clearly disproves this assumption, which, indeed, is in opposition to the prevalent current belief in the existence of schooling effects. Our findings therefore call for reconsideration of the conceptual basis underlying the definition of deviation-IQ scores. If schooling affects intelligence test scores, then aged-based norms "penalize" individuals with less schooling experience. Correcting this bias would require use of school exposure variables in norming.

Second, the considerable effect of schooling found for a variety of general ability tests further blurs the already problematic distinction between the concepts of "intelligence" and "scholastic achievement" (Anastasi, 1984; Cronbach, 1984; Jensen, 1980) and further challenges the notion of "culture-fair" tests. Clearly, the position that "a person's total score [on the Progressive Matrices Test] provides an index of his intellectual capacity whatever his nationality or education" (Raven, Court, & Raven, 1975, p. 1) is no longer tenable.

Third, the revealed between-test variability in the magnitude of the schooling effect suggests that sensitivity to schooling may serve as a dimension for distinguishing between intelligence tests. This dimension corresponds roughly to Cattell's (1963, 1971) and Horn's (1970, 1978) crystallized ability factor and to Cronbach's (1984) spectrum of general abilities. The tests with the lowest schooling effects (figure series, figure analogies, and figure classification) are also the least "crystallized" ones—that is, those that involve the lowest degree of "direct training" (Cronbach, 1984, p. 253)—while those with the highest schooling effects (vocabulary, sentence completion, word arithmetic problems) are the most crystallized. On the other hand, our results do not support Cattell's and Horn's predictions concerning the effects of age and schooling on the development of fluid ability. According to these authors, this ability develops during childhood by means of "incidental" learning (Horn, 1978) independently of schooling. Hence, the effect of schooling on "pure" tests of this ability should be nil and the increase in test scores should be entirely attributable to the age factor. Our results are inconsistent with this prediction: not only is the schooling effect found for most of the six "fluid" tests (tests 2, 4, 5, 7, 8, and 10 in Table 4) considerably greater than zero, but for four of them this effect is equal to or larger than the corresponding age effect.

Fourth, though this study did not specifically include curriculum-related achievement tests, which are at the highest level of crystallization, the results obtained here allow us to draw certain conclusions with regard to these tests. More specifically, the revealed effect of schooling on less crystallized tests can be considered as a lower bound for the effect of schooling on achievement test scores. These results should help counter the skepticism prevailing in some quarters concerning the absolute effect of schooling on children's scholastic achievement (see Coleman, Hoffer, & Kilgore, 1982, and Madaus et al., 1980, for more detailed presentations of this position). They suggest that school attendance does indeed make a difference.

A final remark is in order concerning the specific causes underlying the effects of schooling. While the results of this study clearly point to the considerable effect schooling has on scores obtained on a variety of general ability tests, they do not illuminate the causes of this effect. Obviously, the increase in mean test scores following additional schooling may be due to the similarity between tests and school activities in terms of content or underlying cognitive strategies. However, one cannot readily dismiss the possibility that this increase may also reflect the effects of seemingly "irrelevant" factors, such as test-wiseness. Moreover, the underlying causation may vary both between tests and school levels. Further work is needed in order to elucidate these issues.

References

Anastasi, A. (1984). Aptitude and achievement tests: The curious case of the indestructible strawperson. In B. S. Plake, *Social and technical issues in testing: Implications for test construction and usage* (pp. 129–140), Hillsdale, NJ: Erlbaum.

Anastasi, A. (1986). Intelligence as a quality of behavior. In R. S. Sternberg & D. K. Detterman (Eds.), *What is intelligence?* (pp. 19–22). Norwood, NJ: Ablex.

Angoff, W. H. (1984). *Scales, norms and equivalent scores*. Princeton, NJ: Educational Testing Service.

Binet, A., & Simon, T. (1916). *The development of intelligence in children* (E. S. Kite, Trans.). Baltimore: Williams & Wilkins.

Cahan, S., & Davis, D. (1987). A between-grade-levels approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal*, **24**, 1–13.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, **54**, 1–22.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton-Mifflin.

Cattell, R. B., & Cattell, A. K. S. (1965). *Culture Fair Intelligence Test: Scale 2*. Champaign, IL: Institute for Personality and Ability Testing.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement*. New York: Basic.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design analysis issues for field settings*. Boston: Rand McNally.

Cronbach, L. J. (1984). *Essentials of psychological testing (4th ed.)*. New York: Harper & Row.

deGroot, A. D. (1951). War and the intelligence of youth. *Journal of Abnormal and Social Psychology*, **46**, 596–597.

Flanagan, J. C. (1951). Units, scores and norms. In

E. F. Lindquist (Ed.), *Educational measurement* (pp. 695–763). Washington, DC: American Council on Education.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, **39**, 93–104.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, **5**(10), 3–8.

Harnqvist, K. (1968). Relative changes in intelligence from 13 to 18. *Scandinavian Journal of Psychology*, **9**, 50–53.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Horn, J. L. (1970). Organization of data on life-span development of human abilities. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 423–466). New York: Academic Press.

Horn, J. L. (1978). Human ability systems. In P. B. Baltes (Ed.), *Life-span development and behavior* (Vol. **1**, pp. 211–255). New York: Academic Press.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Madaus, G. F., Airasian, P. W., & Kellaghan, R. (1980). *School effectiveness*. New York: McGraw-Hill.

Ortar, G., & Shachor, A. (1980). *MILTA: A battery of tests for ages 9 through 18*. Jerusalem: Ministry of Education and Culture (Hebrew).

Raven, J. C. (1983). *The Standard Progressive Matrices, 1938–83*. New York: Psychological Corp.

Raven, J. C., Court, J. H., & Raven, J. (1975). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: Lewis.

Reynolds, C. R. (1982). Methods for detecting construct and predictive bias. In R. A. Berg (Ed.), *Handbook of methods for selecting test bias*. Baltimore: Johns Hopkins University Press.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement*, 2d ed. Washington, DC: American Council on Education.

Sternberg, R. J., & Powell, J. S. (1983). The development of intelligence. In J. H. Flavell & E. M. Markman (Eds.), P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 3. Cognitive development* (pp. 341–419). New York: Wiley.

Thorndike, R. L., & Hagen, E. (1971). *Cognitive Abilities Test, Form 1, Levels A–H, Grades 3–12*. Boston: Houghton Mifflin.

Wohlwill, J. F. (1980). Cognitive development in childhood. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development* (pp. 359–444). Cambridge, MA: Harvard University Press.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. New York: Psychological Corp.