Psychological Bulletin
1981, Vol. 90, No. 1, 153-178

0033-2909/81/9001-0153$00.75

Cognitive Development in Retarded and Nonretarded Persons: Piagetian Tests of the Similar Structure Hypothesis

John R. Weisz and Keith Owen Yeates
University of North Carolina at Chapel Hill

Proponents of "developmental" and "difference" theories of mental retardation have long debated the similar structure hypothesis. It holds that when nonorganically impaired retarded and nonretarded persons are similar in developmental level (operationally defined as mental age), they are also similar in the processes and concepts by which they reason. The developmental position endorses this hypothesis; the difference position opposes it. The hypothesis and the debate over its validity have been strongly influenced by Piaget's cognitive developmental theory. In this article we survey 30 studies involving 104 separate tests of the hypothesis with Piagetian conceptual measures. We distinguish between studies that did and did not meet an important procedural requirement: screening organically impaired subjects from their mentally retarded samples. Studies that did not meet this requirement yielded findings inconsistent with the similar structure hypothesis; retarded groups were significantly inferior to nonretarded groups matched on mental age. Studies that did meet the requirement, however, yielded findings supporting the hypothesis. This pattern of findings carries significant implications for the developmental versus difference controversy and for diagnosis and training of the mentally retarded.

When a mentally retarded child and a younger nonretarded child happen to be at the same level of intellectual development, say at the same mental age, how similar are the two children in the processes by which they reason? For a decade and a half this has been a central issue in an often heated debate. The debate has far-reaching implications for the study of both mental retardation specifically and cognitive develpment generally; it is called the *developmental versus difference controversy*.

On one side of the controversy is the developmental position, advanced by Zigler (1969) and elaborated by Weisz, Yeates, and Zigler (in press). This position applies only to people not suffering from organic impairment. It holds that retarded and nonretarded people pass through cognitive developmental stages in an identical order. These stages are usually defined as those described by Piaget (e.g., 1964, 1970). Retarded people are said to traverse these stages more slowly and attain a lower upper limit than nonretarded people. But when the two groups are equated for level of cognitive development, which is most often operationally defined as psychometric mental age (MA), they are not expected to differ in the cognitive processes by which they reason.

On the other side of the controversy are theorists who hold what Zigler (1969) has labeled the difference position. One tenet of this position is that the cognitive development of the retarded person differs from that of the nonretarded person in ways other than mere differences in developmental rate and ceiling. Even when retarded and nonretarded people are equated for level of development, they are expected to differ in the cognitive processes they use in reasoning. Milgram (1973), for example, has argued that the

The review was facilitated by support from National Institute of Mental Health Research Grant 1 RO3 MH34652 01 to John R. Weisz and National Institute of Health and Human Development Training Grant 1 T32 HDO7178 to Keith Owen Yeates.

The material discussed in this article is also reviewed in the forthcoming book, *Mental Retardation: The Developmental-Difference Controversy*, edited by Edward Zigler and David Balla.

We are grateful to Thomas Achenbach and David MacPhee for their helpful comments on an earlier version of this article.

Requests for reprints should be sent to John R. Weisz, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina 27514.

cognitive stages of retarded individuals are structurally different from those of the nonretarded. The retarded person's stages, he argued, are likely to contain traces of more primitive developmental stages and are apt to give way to a regression to those earlier stages. Consequently, he maintained that retarded persons employ less advanced reasoning than MA-matched nonretarded persons.

By its emphasis on processes of reasoning, the developmental versus difference controversy has helped to stimulate a significant shift in the focus of comparative research. There has been a reduced emphasis on such end products of learning and problem solving as trials to criterion on experimental tasks. Investigators have been increasingly attentive to the dynamics underlying such learning and problem solving (see e.g., studies of hypothesis behavior by Weisz, 1977; Weisz & Achenbach, 1975). The emphasis on process has also stimulated a burgeoning interest in comparative research on the concepts people use in reasoning. These concepts, as described by Piaget (e.g., 1970), are the very implements of thought, the tools by which reasoning occurs.

Along with heightened interest in Piagetian theory has come an awareness that the developmental versus difference debate really involves two hypotheses. Either may be true or false independent of the other. One is the hypothesis that retarded and nonretarded individuals pass through cognitive developmental stages in the same order. This has been labeled the *similar sequence* hypothesis. Evidence germane to it was examined by Weisz & Zigler (1979), and most of the evidence supported the hypothesis. The second hypothesis addressed the question with which we began this article, and it is the central focus of this review. We describe the hypothesis and discuss its importance in detail.

Similar Structure Hypothesis and Piagetian Research

As noted earlier, the developmental position holds that retarded and nonretarded people who are matched for level of development are also alike with respect to their formal processes of reasoning. We label this tenet the similar structure hypothesis. We use this label because the hypothesis is based on the view that such carefully matched retarded and nonretarded people are similar in cognitive structure as described by Piaget (see Zigler, 1969, pp. 537–541). In the Piagetian literature the term *cognitive structure* is used to represent the way intellect is organized (e.g., the logico-mathematical forms that underlie particular patterns of reasoning). Cognitive structure cannot be directly observed. Instead it is inferred from various overt manifestations of reasoning. These manifestations constitute operational definitions of cognitive structure. Typically they take the form of Piagetian measures of the concepts children use in reasoning and solving problems. Thus, studies using such measures form the pool of evidence used in the present review.

The similar structure hypothesis holds that when these Piagetian measures are used with retarded and nonretarded groups of similar MA, and when appropriate experimental control has been exercised (as described below), the groups will not differ reliably. In sharp contrast, Milgram (1973) argues that the retarded will prove inferior to their nonretarded MA peers in the conceptual processes required to perform Piagetian tasks. He argued that IQ is an index of problem-solving ability, and he cites Inhelder's (1968) view that the retarded child's reasoning shows "viscosity, oscillation," and regression to immature levels. Building on this rationale, Milgram (1973) predicted that "given tasks without a ceiling or floor effect for the general MA of the subjects being used, the retarded will dependably demonstrate an equal-MA deficit" (p. 209).

In arguing that the retarded will be inferior to their nonretarded peers, Milgram adopted a view that is typical of most difference positions. A rather unconventional difference position has been taken by Kohlberg (1968). He believes that the kind of conceptual development described by Piaget requires massive doses of "general experience." Consequently, Kohlberg reasons that the retarded should be conceptually more advanced than their nonretarded MA peers because they have lived longer and thus have acquired a more extensive base of experience.

For two reasons, then, evidence from the Piagetian tradition seems ideally suited to a test of the similar structure hypothesis. First, the hypothesis itself was stimulated by attention to the kinds of underlying processes Piagetian conceptual measures are designed to tap. These measures are quite diverse; they cover a broad spectrum of the child's experience with the world. Second, it is specifically within the Piagetian domain that an unusual triangular theoretical conflict has taken shape. There are three possible outcomes of experimentation comparing MA-matched retarded and nonretarded groups. Each outcome can add to our knowledge by supporting one of three incompatible views—the developmental position and its similar structure hypothesis (Zigler, 1969), the conventional difference position (Milgram, 1973), or an unconventional difference position (Kohlberg, 1968). This three-way theoretical conflict presents an opportunity for the kind of strong inference rarely seen in psychological research. In the present article we attempt to take advantage of this unusual opportunity. First, certain caveats are in order.

Methodological Issues: Ground Rules for a Fair Test

Matching for Developmental Level

The similar structure hypothesis predicts similar reasoning in children who are equated for developmental level. The most widely employed operational definition of developmental level in research on the developmental–difference controversy is the psychometric MA. Certainly, MA is at best a rough index; it could possibly be improved upon by a collection of measures derived from developmental theory. Nonetheless, the MA, as obtained on such standardized instruments as the Stanford-Binet Intelligence Test, is a broad-based measure that taps a wide range of cognitive activities. Moreover, nonorganically impaired mentally retarded and nonretarded children who are matched for total Stanford-Binet MA (both short form and standard) are evidently similar in the patterns of items they pass (Achenbach, 1970, 1971). Although the MA may be a less than ideal measure of developmental level, it appears to be an adequate matching device for tests of the similar structure hypothesis.

The Question of Organic Impairment

The developmental position generally, and the similar structure hypothesis in particular, both apply only to people not afflicted with specific genetic or other physiological defects. Thus the theory is not intended to apply to people whose retardation is linked to such causes as Down's syndrome or brain injury. Some critics of the developmental position (Ellis, 1969; Milgram, 1969, 1973; Spitz, in press) have maintained that these kinds of organic impairments do not influence behavior significantly. Consequently, they have argued that it is not necessary to exclude organically impaired persons when structuring tests of the developmental position. Ellis, (1969) for example, insisted that only "rarely have behavioral differences characterized different etiological groups" (1969, p. 561). Reviews touching on this question reveal rather mixed evidence (e.g., Blount, 1968). It should be noted, however, that a number of investigators (e.g., Cruikshank, 1967; Reitan, 1973) have devoted much of their lifework to the study of idiosyncratic behavior distinguishing organically impaired individuals from those with intact nervous systems. Furthermore, there is some evidence that organic impairment may affect performance on the specific kinds of reasoning and problem-solving tasks most often used in research on the developmental versus difference controversy; effects have been found in mentally retarded groups (see Balla, Styfco, & Zigler, 1971; Balla & Zigler, 1964; Harter, Brown, & Zigler, 1971) and in nonretarded groups (Elkind, Koegler, Go, & Van Doornick, 1965).

In summary, some but not all studies have revealed performance anomalies associated with organic impairment. The developmental position specifically excludes organically impaired individuals from its predictions. Furthermore, difference theories (with the possible exception of Luria's, 1963, position) can be adequately tested with samples that exclude organically impaired subjects. Thus the clearest tests of the similar structure hypothesis are those that exclude organically abnormal cases. This does not mean that studies failing to satisfy this condition have been excluded from the present review. It does mean, however, that such studies must be regarded as yielding less than ideal evidence. At the end of our review, we try to determine whether studies in which organically impaired individuals were excluded showed a different pattern of findings than studies in which organically impaired individuals were not excluded.

Role of Noncognitive Factors

Another consideration is that the behavior of retarded people reflects more than their formal cognitive processes. There is general agreement that a relative deficiency in cognitive functioning is the essential defining feature of mental retardation. However, noncognitive factors distinguishing retarded from nonretarded persons may have obscured the precise nature of this cognitive deficiency. Over the past 2 decades the work of the Peabody group (see Cromwell, 1963), Gordon and O'Connor (reviewed by Heber, 1964), Stevenson and his colleagues (Stevenson, 1965), and the Yale group (see Zigler, 1971), has revealed motivational characteristics of the retarded that undermine their experimental task performance. For example, numerous studies by the Yale group (see Zigler, 1971) have exposed the impact of such factors as institutionalization, heightened desire for attention and praise, wariness of adults, idiosyncratic reward preferences, tendency to anticipate failure, and willingness to settle for relatively low levels of success. There is also a growing body of evidence on learned helplessness among the retarded; it indicates that retarded children, particularly older ones, are less likely than nonretarded children of similar MA to ascribe their failures to controllable factors and are less likely to persist in the face of failure and negative feedback (see Gibson, 1980; Weisz, 1979, 1981, in press-b).

Given such findings on motivation and helplessness, it is clear that for tests of the similar structure hypothesis to be fair, retarded and nonretarded subjects should be (a) equally motivated to perform at their best and (b) equally confident that success is possible. This standard reflects a simple principle: One cannot conclusively attribute a group difference on a dependent variable to one known difference in subject characteristics (e.g., IQ) when the subjects also differ on other factors that affect performance on that variable. The importance of this principle is underscored in the writings of Cole and Bruner (1971), Labov (1970), and Tulkin and Konner (1973) on the dangers of attempting to infer competence directly from performance.

Having stressed the importance of controlling noncognitive factors, we must note that few of the studies we review show explicit efforts to exercise such control. At this point in the evolution of the developmental–difference controversy, however, we believe it is important to take the proverbial bull by the horns and use such data as are available in the literature. We have therefore included all the studies we could find that we considered relevant, whether or not they included efforts to control motivation and expectancy; however, we have attempted to provide sufficient details of each study to enable the reader to judge whether noncognitive factors may have influenced the findings.

Variability and the Reliability of "No-Difference" Findings

The reader should be sensitive to one other aspect of the studies reviewed. The similar structure hypothesis is fundamentally a null hypothesis. That is, a conclusion that the hypothesis is supported can result only from findings of no significant difference between MA-matched groups differing in IQ. Yet no-difference findings have traditionally been regarded by social scientists as inconclusive. We discuss the implications of this paradox after reviewing the data. During the review, however, we attempt to shed light on the question of whether the no-difference findings might have resulted from excessive variability in the data. Such an interpretation of a no-difference finding loses considerable credibility when other analyses of the same subjects yield statistically significant differences. We therefore indicate, with each no-difference finding reported, whether other differences among the same subjects attained significance.

Having registered these caveats, we now proceed to the evidence on the similar structure hypothesis, drawn from several different conceptual domains.[1]

Tests of the Similar Structure Hypothesis

Moral Judgment

Retarded people have often been labeled immoral by the nonretarded in our society (see accounts in Kamin, 1975; Sarason & Doris, 1969; Zigler & Harter, 1969). Consequently, research comparing the moral judgment of retarded and nonretarded persons should be of particular interest. Our literature search yielded two relevant experiments, both carefully designed. In the first, Taylor and Achenbach (1975) compared the judgments of MA-matched retarded and nonretarded subjects at three MA levels: 6½, 8, and 9½ years. Retarded and nonretarded groups had mean IQs of 75 and 111, and mean chronological ages (CAs) of 11.4 and 6.9 years. Subjects judged what various protagonists should do in stories involving moral dilemmas. The dilemmas concerned promises, stealing, punishment, and the value of life. The youngsters' rationales for their judgments were scored according to Kohlberg's (1969) stages of moral reasoning. All subjects were from public schools: records were used to select only children showing no indication of organic impairment. The retarded and nonretarded groups did not differ significantly in their level of moral judgment. There was a significant effect of MA level on moral judgment, however. This suggests that the no-difference finding did not result from excessive variability in the data. The Taylor–Achenbach study included Piagetian measures in other conceptual domains. These are discussed later in this review.

The second study was designed partly as a sequel to the Taylor–Achenbach experiment. Kahn (1976) compared moral judgments of three MA-matched groups in a five-story interview based on Kohlberg's procedures. His subjects were all public school children: (a) nonretarded subjects averaging 101 in IQ, 7.3 years in MA, and 6.9 years in CA; (b) mildly retarded subjects averaging 66 in IQ, 7.5 years in MA, and 10.7 years in CA, all screened to eliminate organic cases; and (c) moderately retarded subjects averaging 45 in IQ, 7.3 years in MA, and 18.2 years in CAs, with etiologies including Down's syndrome, rubella, congenital syphilis, and other organic abnormalities. Consistent with the similar structure hypothesis, the first two groups did not differ in level of moral judgment; in fact, the second group was not significantly different from either of the other groups. However, the group suffering from organic impairment was significantly inferior to the nonretarded group in moral judgment. Kahn's conclusion about this pattern of findings directly addresses the etiological issue we discussed earlier: "Similarity of performance by [MA-matched] retarded and nonretarded children does not extend to retarded children with etiologies other than cultural-familial retardation" (1976, p. 213). Other measures used by Kahn are discussed below.

Role Taking

One conceptual skill that many consider related to moral judgment is role taking. The Taylor and Achenbach (1975) investigation included two role-taking tasks. In the first, the subject's level of awareness of the experimenter's thought processes was inferred from behavior on successive games of tic-tac-

---

[1] The evidence that is reviewed with regard to the similar structure hypothesis unfortunately does not include any of the data from the Temple longitudinal study (Stephens, et al., 1974), described in the review of the similar sequence hypothesis (Weisz & Zigler, 1979). Although the Temple data are potentially applicable to the similar structure hypothesis, the published analyses of the data (Mahaney & Stephens, 1974; Stephens, 1972; Stephens et al., 1974) are not. Retarded and nonretarded groups in the study are not matched for MA, and when analysis of covariance is employed (see Stephens & McLaughlin, 1974) in contrasting the performance of retarded and nonretarded subjects, it is used to control for both MA and CA. This analysis, which in principle also controls for group differences in IQ, is not appropriate for a test of the similar structure hypothesis (see Weisz, 1976) and poses other statistical problems as well (see Kappauf, 1976).

toe. The second was a guessing game designed to tap higher levels of role-taking skill (adapted from Flavell, Botkin, Fry, Wright, & Jarvis, 1968). In this game the child tried to fool an observer who had to guess which of two boxes the child had taken pennies from. When responses to both tasks were scored according to Kuhn's (Note 1) role-taking criteria, the retarded and nonretarded groups did not differ. There was a significant MA effect on the simpler role-taking task and a significant sex effect (boys superior) on the more advanced task.

A large-scale study by DeVries (1970, 1973a, 1973b) included a role-taking task in which the child tried to fool an experimenter who guessed which hand the child had hidden a penny in; at other times the child did the guessing. Intellectually average children with a mean IQ of 105, mean MA of 6.8 years, and mean CA of 6.5 years were included. These youngsters were compared with a retarded group with a mean IQ of 72, mean MA of 6.5 years, and mean CA of 8.8 years. Retarded children were evidently not screened for organic impairment. In contrast to the findings of Taylor and Achenbach (1975), DeVries found retarded children to be significantly inferior to the nonretarded in role-taking performance. DeVries used other measures that we discuss later.

Perspective Taking and Perceptual–Spatial Concepts

A number of problems outside the social realm per se involve reasoning that is conceptually similar to role taking.

Perspective-taking (spatial egocentrism). One problem is that of identifying how a perceptual array would look from a perspective other than one's own. Rubin and Orr (1974) compared MA-matched retarded and nonretarded children on two such tasks; one task involved identifying and the other, constructing the perspective of another person on a stimulus array. Both groups of children were from public elementary schools. Their IQs averaged 73 and 105, their MAs averaged 7.7 years and 8.5 years, and their CA means were 11.9 and 7.9 years, respectively. There was apparently no screening of organic cases. Nonetheless, the two groups did not differ reliably on either task; however, the retarded group was significantly inferior to a CA-matched group on both tasks.

Perceptual decentering. Smith (1977b) assessed a related skill using a modification of Elkind, Koegler, and Go's (1964) Picture Integration Test. The test consisted of several complex pictures (e.g., a face composed of a bat, balls, acorns, and a drum). To assess the child's ability to decenter, the experimenter first asked what the picture looked like and then whether the child saw anything else. Responses were scored for the number of responses involving parts, the number involving whole "gestalts," and the number of part–whole integrations (the most advanced response developmentally). Again the subjects were MA-matched retarded and nonretarded public school children. The two groups had mean IQs of 69 and 115, the MA means were 8.6 years and 9.1 years, and CA means were 12.5 years and 8.5 years, respectively. The retarded sample was evidently not screened for organically impaired children. The two groups did not differ significantly on the decentering task; however, both groups had significantly more part–whole integrations than did a retarded group matched to the nonretarded group for CA.

Relative thinking. Another related problem is that of relative thinking as measured by Piaget's (1928) "brother-sister" and "right-life" tests. The first test taps children's grasp of the relation between being a sibling and having a sibling. A sample item: George has three brothers, Paul, Henry, and Charles. How many brothers does Paul have? How many brothers are there in the family? On the right–left test, children point to their own right and left hands, then to the experimenter's. McManis (1969f) gave both tests to 15 retarded and 15 nonretarded youngsters at each yearly MA level from 5 to 10 years. The IQs averaged 58 for retarded subjects and 100 for the nonretarded; CAs ranged from 7.7 to 21.2 years in the retarded group and from 5.3 to 10.8 years in the nonretarded (CA means were not reported). The source of subjects (public schools vs. residential institutions) was not specified. McManis (1969f) reported no significance tests. He indicated, however, that on both tasks the pattern of results was similar: as level of abstractness in the test questions increased, retarded subjects showed increasing MA deficits (i.e., correctly answered questions at later MA levels than nonretarded subjects). Given the well-established linkage between brain damage and difficulty with abstraction, it is important to note that McManis's (1969f) retarded sample included substantial percentages suffering from organic impairment: 62% at MA 5, 43% at MA 6, 43% at MA 7, 23% at MA 8, 8% at MA 9, and 46% at MA 10.

Lane and Kinder (1939) used the same two Piagetian tasks to assess relative thinking among five groups of institutionalized retarded subjects of unspecified etiology. Two of the groups differed in IQ but were matched for MA. A group with a mean IQ of 51 averaged 8.3 years in MA and 16.3 years in CA. A group with a mean IQ of 64 averaged 8.1 years in MA and 12.1 years in CA. No significance tests were reported anywhere in the study. However, the two MA-matched groups were quite similar, both in the gross percentage of questions answered correctly and in the percentage of subjects answering each individual question correctly. Other groups in the sample differed markedly in both respects. Lane and Kinder (1939) note that "the records of the two groups with the same mental age are thus found to be strikingly similar, whereas comparisons on the basis of chronological age or IQ show marked differences in performance" (p. 113).

In a third study of relativism, Prothro (1943) administered the right–left test to MA-matched institutionalized retarded and public nursery and elementary school nonretarded subjects. The nonretarded sample had a mean IQ of 113, a mean MA of 6.2 years, and a mean CA of 5.6 years. The retarded sample had a mean IQ of 39, a mean MA of 6.1 years, and a mean CA of 27.3 years; only retarded individuals diagnosed "familial" or "undifferentiated" were included; cases involving "motor defects" or epilespy were excluded. It is not clear what etiologies may have characterized members of the undifferentiated group. It is possible, however, that some of them were organically impaired, since an IQ of 50 is the approximate lower limit for the familial retarded (see Zigler, 1967), and Prothro's group mean was 39. Performance of the nonretarded children on the right–left test, though at a very low level, was significantly superior to that of the retarded. We return to the Prothro study when we discuss animism, causality, and quantitative concepts.

The Devries (1970, 1973a, 1973b) and Taylor and Achenbach (1975) studies also assessed relative thinking. Recall that Devries's study compared retarded children averaging 71 in IQ and 6.5 years in MA with nonretarded children averaging 105 in IQ and 6.8 years in MA; the retarded group was apparently not screened for organic impairment. Taylor and Achenbach did screen out organically impaired cases. Their sample included retarded and nonretarded children with mean IQs of 75 and 111, respectively; the groups were MA-matched at three MA levels: 6½, 8, and 9½ years. Devries found no significant difference between retarded and nonretarded subjects on either the brother–sister or right–left test, but the brother–sister test did differentiate significantly between high-IQ and average-IQ children of the same CA. Similary, Taylor and Achenbach (1975) found that whereas performance on the two tests improved significantly with MA, there were no significant differences between retarded and nonretarded children matched for MA.

Logical contradication. As a child moves into the concrete operational stage of cognitive development, an important transition takes place. Perceptually dominated thought gives way to logically guided thought. That is, cognition begins to be less "stimulus bound." Smith (1977a) investigated the relative progress of retarded and nonretarded subjects into logically guided thought. He compared the performance of 30 nonretarded and 30 retarded school children matched for MA on a task involving logical contradiction. The IQ, MA, and CA means were not reported, and there was no indication that the children were screened for organic impairment. The task was a modification of the Jastrow ("ring segment") Illusion. In this illusion, the child is first asked to pick the "big" ring segment. The position of the two ring segments is then reversed, and the subject is asked to pick the "little" ring segment; when position is changed, the ring first identified as big by the child appears to be little. Smith (1977a) and an independent observer rated reactions to the illusion, looking for evidence of surprise at the contradiction between logic and perception. They found that the MA-matched retarded and nonretarded children did not differ significantly in frequency of surprise reactions. However, retarded children showed surprise significantly less often than a CA-matched nonretarded group.

Conceptions of the World and Physical Causality

Animism. One of Piaget's earliest interests concerned children's interpretations of naturalistic phenomena. Of the many that Piaget investigated, one of the first subjected to American comparative research was the child's concept of life. Russell, Dennis, and Ash (1940), administered a standardized interview in which subjects were asked to classify a series of objects as living or nonliving and to justify each classification. The retarded subjects were inmates in two institutions for the "feeble-minded"; they were not described with respect to etiology or IQ. A nonretarded contrast group (no IQ or institutional status reported) was matched to the retarded group for MA range at two MA levels (i.e., 6–7 and 8–9 years). Neither MA nor CA means were reported for retarded or nonretarded subjects. At the lower MA level the two groups did not differ significantly in animism stage level. At the higher MA level, however, the retarded subjects had significantly more advanced life concepts than their nonretarded MA peers. Following up on this finding, Russell et al. compared that portion of the retarded sample who were under 21 years of age with their older retarded MA peers; the older group showed significantly more advanced life concepts. Russell et al. concluded that although MA imposes restrictive limits on the effects of CA, experience does seem to enhance the development of the life concept over and above the effects of MA alone. This conclusion and the finding on which it is based represent the first support we have found for Kohlberg's (1968) unconventional difference position on retardation.

However, when Prothro (1943) used the Russell et al. procedure he failed to find support for Kohlberg's (1968) position. In Prothro's study, described earlier, a retarded group screened for motor defects and epilepsy averaged 39 in IQ and 6.1 years in MA. The nonretarded comparison group averaged 113 in IQ and 5.6 years in CA. The two groups did not differ significantly in life concepts.

Granich (1940) used an interview similar to but more diversified than that used by Russell et al. and by Prothro. Granich's sample included public school boys of average and retarded intelligence matched for MA range (7.2–8.4 years). The nonretarded boys averaged 102 in IQ, 7.8 years in MA, and 7.6 years in CA. The retarded boys averaged 56 in IQ, 7.7 years in MA, and 13.9 years in CA. The etiologies for the retarded sample were not reported. The two subject groups were very similar in frequency of attribution of life (or traits of living things) to animate and inanimate objects; both groups were markedly inferior to a group of normal children averaging 13.6 years in MA. No significance tests were reported, however. The Granich interview included questions on several other conceptual domains that are discussed later in this review.

More recently, Smeets (1973, 1974) studied children's attributions of life and specific life traits (e.g., dying, growing, feeling) to animate and inanimate objects. In his first study, children's attributions were followed by the question, What makes you think so? The sample included retarded and nonretarded subjects matched for MA range. The source of subjects was not specified. Group means were not reported, but the nonretarded group ranged from 90 to 110 IQ, from 5.8 to 7.2 years in MA, and from 5.8 to 7.2 years in CA. The retarded group ranged from 50 to 65 in IQ, from 5.8 to 7.2 years in MA, and from 11.0 to 12.0 years in CA. In addition, an older nonretarded group (IQs 90–110) were matched to the retarded group for CA range (11.0–12.0 years). Errors were most frequent among the young normal subjects and least frequent among the older normal subjects, but no tests of significance were reported. However, a probability analysis of the distribution of errors indicated that the errors made by the retarded and younger normal groups were largely due to guessing; the errors made by the older normal group were less frequent and less capricious. In a second report, apparently based on the same experiment, Smeets (1974) presented significance tests for differences among the groups. In the attribution of life to animate objects, the three groups differed significantly overall, but the retarded and MA-matched nonretarded did not. In the attribution of life to inanimate objects and in the attribution of life traits to animate objects, the three groups did not differ significantly. Finally, attribution of life traits to inanimate objects was significantly less frequent among the older normal than the younger normal group, but neither group differed significantly from the retarded.

Artificialism. In Piaget's (1929) book, *The Child's Conception of the World*, he included among the child's "prelogical" beliefs not only animism but artificialism—the tendency to explain natural phenomena as resulting from deliberate activity of man or anthropomorphic beings. The study by Granich (1940), described previously, included a number of questions about the origin of natural phenoma (e.g., flower seeds). Granich found that responses indicative of artificialism were virtually nonexistent among the 13- and 14–year-old nonretarded boys in his sample. Such responses were slightly less frequent among the sample of retarded boys than among their normal MA peers. The difference between thes latter two groups was not tested for significance, but it is very small and would almost certainly have been nonsignificant.

Realism and the dream concept. A third prelogical belief attributed by Piaget (1929) to the young child is realism. One reflection of this belief is the inability to distinguish between external objects and one's thoughts and dreams. Granich's (1940) interviews included questions designed to tap realism; for example, one asked whether dreams actually appear during the night on the ceiling of the bedroom. Granich found that 13- and 14-year-old boys showed virtually no realism. The retarded group showed slightly fewer such responses than their nonretarded MA peers; this difference would almost certainly have been nonsignificant had Granich tested it.

DeVries's (1970, 1973a, 1973b) investigation also included an assessment of children's dream concepts. In this portion of her study, average-IQ children differed significantly from their CA-peers of high IQ, but the MA-matched low- and average-IQ children did not differ significantly.

Magic. In a related feature of DeVries's study, children's notions with regard to magic (see Piaget, 1930) were assessed. Here again, the CA-matched average- and high-IQ children did differ significantly, but the MA-matched average- and low-IQ subjects did not.

Causality. In *Judgment and Reasoning in the Child*, Piaget (1928) discussed how the child's developing ability to understand causal relations is reflected in a capacity to use the word *because* correctly. In Prothro's (1943) comparative study, the subjects were all asked to complete five sentence beginnings that ended in the word *because.* (e.g., "The boy hurt his leg because . . ."). Responses were scored for whether they actually described a causal relation. Prothro's retarded sample was significantly less adept at this task than his MA-matched group of nonretarded children.

Quantitative Concepts (Other Than Conservation)

Classification and class inclusion. Grouping objects into classes is one of the simplest quantitative behaviors. Apparently the earliest comparative study of classification that bears on the similar structure hypothesis was the Prothro (1943) investigation. Prothro used three tasks that called for the child to sort objects into groups of things "that are just alike" or "that go together." The tasks involved sortings beads, pencils, and pasteboard shapes. Classification on the basis of color and form were both possible in each task. With scores summed across the three tasks, Prothro's institutionalized retarded subjects were significantly inferior to nonretarded children of similar MA. Since this task involves abstraction, there is some question as to whether Prothro's findings may have been influenced by the possible inclusion of organic cases among his retarded subjects.

At first glance, the Taylor and Achenbach (1975) study appears to provide a useful contrast to the Prothro experiment, since Taylor and Achenbach excluded all cases with signs of organicity, and their procedure included a classification task quite similar to those used by Prothro. Their finding of no-difference between MA-matched retarded and nonretarded subjects must be regarded as not very meaningful, however; the fact that all of their subjects succeeded at the task indicates that it was too simple to reveal group differences.

Finally, the DeVries (1970, 1973a, 1973b) investigation included a classification ("sorting") and a class inclusion problem, though neither problem is described in detail. On neither problem did DeVries's retarded and average-IQ children of similar MA differ significantly. This may reflect a floor effect on the sorting task, which did not show a CA effect either. But the no-difference finding with the class inclusion task appears to be more meaningful: this task did show a highly significant effect of CA ($p < .0001$).

Additive composition. A concept closely linked to classification is additive composition. Piaget (1952) described a three-stage process in the development of the ability to additively compose classes. Children in the first stage cannot compare subclasses with whole classes but can compare separate subclasses; during the second stage, children identify subclasses with the whole; finally, at the third stage, children have an abstract conception of subclasses within wholes. To illustrate, Piaget showed children an array of brown beads and white beads, all of which were wooden. When asked whether there were more brown beads or wooden beads, Stage 1 children answered by comparing the brown subclass to the white subclass. Stage 2 children identified the brown subclass with the whole class "wooden," and answered "the same." Only in Stage 3 did children recognize that there were more beads in the whole class "wooden" than in the subclass "brown." McManis (1968) presented a similar task to 90 nonretarded elementary school children and 90 institutionalized retarded children, matched at six MA levels (5, 6, 7, 8, 9, and 10–11 years). The retarded subjects were evidently not screened for organic impairment. Group means were not reported, but the nonretarded group ranged from 85 to 115 in IQ, from 5.3 to 11.9 years in MA, and from 5.3 to 10.8 years in CA. Retarded subjects ranged from 47 to 73 in IQ, from 5.0 to 11.8 years in MA, and from 7.7 to 21.1 years in CA. McManis (1968) found a significant increase in the proportion of Stage 3 responses and a decrease in proportion of Stage 1 responses across MA levels for both retarded and nonretarded children. (No Stage 2 responses were obtained.) At each of the five MA levels, the retarded subjects showed more Stage 3 responses than did the MA-matched nonretarded group. This difference favoring the nonretarded was not statistically significant, however.

Concept of probability. Children's concepts of probability were studied by Stevenson, Hale, Klein, and Miller (1968). The investigators employed filmed instructions and a test booklet with intact school classes, each seen as a group. To introduce the probability questions, the film noted that in a group of boys pictured in the test booklet, some had dark hair and some light, whereas some wore glasses and some did not. Subjects were asked to infer which of two characteristics was more probable, given certain information about base rates in the target population (e.g., "I am thinking of a certain boy in the group who wears glasses. What color do you think his hair is, light or dark?" [Stevenson et al., 1968, p. 19]). In a second task, the youngsters were told how may pegs of various colors were inside a certain box and then were asked to predict the outcome of blind drawings from the box. The original sample included high-IQ children from a university laboratory school, average-IQ children from a public junior high school, and low-IQ children from public school classes for the educable retarded. Mean IQs of the three groups were approximately 120, 100, and 70, respectively. Retarded subjects were reported to be "primarily" of the familial type, with cases of "gross motor or sensory impairment" excluded. Stevenson et al. (1968) did not attempt to achieve MA matching but they did compare the performance of retarded subjects with a fourth-grade, nonretarded group whose MA was said to be "comparable with that of the retarded *S*s" (p. 29). Mean MAs of the two groups were not reported, nor was any test reported to indicate whether the two groups differed significantly in MA. On both tests of the probability concept, the retarded subjects were significantly inferior to the nonretarded group. The study included other measures, most of which are not relevant to this review. But one, a conservation task, is discussed below.

Comparison of gross, intensive, and extensive quantities. In his analyses of the development of quantitive thinking, Piaget (1952) described three kinds of perceived quantity by which things are compared without actual measurement. The first to develop is *gross quantity*; it is in evidence when children consider only uncoordinated perceptual relations of gross quantitative equality or difference. In the second, *intensive quantity*, children are said to compare quantities by seriating them along more than one dimension (e.g., width and height) concurrently. The third and most advanced was labeled *extensive quantity*; using this type of perceived quantity, children can overrule apparent differences between two equal quantities by imposing equal units of measurement on them. McManis (1969a) used this typology to score children on their use of comparison processes with sticks, colored water, and beads. His retarded subjects were institutionalized; more than one third were organically impaired. Their mean IQ was 58; they averaged 7.9 years in MA and 14.4 years in CA. The nonretarded subjects were from public schools. They averaged 100 in IQ, 8.5 years in MA, and 8.0 in CA. Across all three types of comparison combined, nonretarded subjects significantly outperformed their retarded MA peers. Separate analyses for each task revealed that the two groups performed similarly in comparing gross quantities but that the retarded were significantly inferior in comparing intensive and extensive quantities. To understand this pattern of findings, McManis (1969a) examined the cognitive processes required for comparison of intensive and extensive quantities. This led him to hypothesize that retarded persons may show deficits in seriation relative to their nonretarded MA peers.

Seriation. In a subsequent study McManis (1970) tested this hypothesis. His retarded sample again consisted of institutionalized individuals who apparently had not been screened for organic impairment. Their IQs ranged from 46 to 72, their MAs from 5.4 to 11.9 years, and their CAs from 8.8 to 22.3 years. (Means were not reported.) The nonretarded comparison group were drawn from public elementary schools. They ranged from 85 to 116 in IQ, from 5.3 to 11.9 years in MA, and from 6.0 to 11.8 in CA. To equate the two groups in developmental level, MA-matched retarded–nonretarded subject pairs were formed. Subjects were asked to arrange five sticks in order of increasing length, then to insert four additional sticks of differing lengths into the arrangement in their appropriate position. As predicted earlier by McManis (1969a; 1969b), significantly more nonretarded than retarded subjects succeeded at the task. When the analysis was broken into five different MA levels, McManis found that the retarded–nonretarded difference was only significant for subjects in the 5–7 year MA range. This study also involved measures of transitivity and conservation, which are discussed later.

Unlike the McManis (1970) study, the investigation by Taylor and Achenbach (1975) systematically ruled out retarded children suffering from organic impairment; the Taylor and Achenbach (1975) study also included a measure of logical seriation. It involved the following problem: "There are three brothers—Bob, Joe, and Dick. Bob is taller than Joe, and Joe is shorter than Dick. Who is the shortest brother?" (p. 47). This problem requires the ability to coordinate quantitative relations into a triadic series. Taylor and Achenbach found no significant difference between their MA-matched retarded and nonretarded subjects, but they also found no significant effect of MA ($p = .13$). This raises a question as to whether the task employed may have simply been insensitive to group differences within the 5- to 10-year-old MA range employed.

Transitivity. A transitive inference involves reasoning like the following: If A > B, and B > C, then A > C. Most Piagetians believe such reasoning is closely tied to seriation. We have found five studies of transitivity that have some bearing on the similar structure hypothesis. One is the DeVries (1970, 1973a, 1973b) investigation. Her battery of tasks included a "length transitivity" task. (No other task details were reported.) The task did differentiate significantly between CA-matched high- and average-IQ subjects, but low- and average-IQ subjects of similar MA did not differ significantly.

Gruen (1973) assessed transitivity of length among MA-matched familial retarded and nonretarded children. Both groups were from public schools. The retarded group averaged 64 in IQ, 6.5 years in MA, and 9.6 years in CA. Nonretarded children averaged 102 in IQ, 6.6 years in MA, and 6.5 years in CA. Three sticks of differing length were used, with Stick A longer than B, and B longer than C. Subjects made comparisons between A and B and between B and C; then they were asked to infer the relative sizes of A and C. To prevent solutions based on perceptual acuity alone, sticks A and C were placed inside the arms of a Mueller–Lyer illusion. This made the shorter stick appear to be the longer one. Half of the subjects in each group were given a memory aid to remind them of the results of the A versus B and B versus C comparisons. In the memory-aid condition, retarded subjects made significantly fewer correct transitive judgments than the nonretarded subjects; in the no-memory-aid condition, the two groups did not differ significantly. For nonretarded subjects performance was significantly better in the memory-aid than in the no-memory-aid condition; among retarded subjects the memory aid made no significant difference. Gruen's interpretation was that the nonretarded children showed superior logical inference. That is, when all the information ordinarily stored in memory was concretely available, retarded children did not use this information in the same logical, sequential manner as their nonretarded MA-peers.

The Mueller–Lyer illusion procedure was also used by McManis (1970) to assess transitivity of length among the nonretarded and unscreened retarded samples described earlier. Recall that MAs in both groups ranged from about 5 years to about 12 years, and the IQ ranges were 85–116 and 46–72, respectively. Although no concrete memory aid was employed, subjects were required to recall the results of the original A versus B and B versus C comparisons before being asked to make the A versus C judgment. Nonretarded subjects made significantly more correct transitive inferences than did the MA-matched retarded subjects.

In another comparative investigation, McManis (1969c) assessed transitivity of length and weight. In the length assessment the Mueller–Lyer procedure was used. Transitivity of weight was assessed via a conceptually similar procedure in which the weight of clay balls of different sizes was varied by embedding differing amounts of metal within them. The sample appears to be the same as in the McManis (1969a) study. There the retarded subjects were all institutionalized, and more than one third were organically impaired. They averaged 58 in IQ and 7.9 years in MA. The public school nonretarded contrast group had a mean IQ of 100 and a mean MA of 8.5 years. McManis (1969c) did not report significance tests for the retarded versus MA-matched nonretarded contrast in his sample; however, our own chi-square analyses of his data indicated that the retarded group made significantly fewer correct transitive inferences than the nonretarded, on both the length and the weight (both *ps* $< .01$).

Although the magnitude and consistency of Gruen's (1973) and McManis's (1969c) findings on transitivity are striking, their methodology has come under fire. Lutkus and Trabasso (1974) have expressed two significant concerns, particularly with regard to the Mueller–Lyer transitivity procedure. (See also Bryant & Trabasso, 1971.) First, they note that the Mueller–Lyer illusion produces a gratuitous perceptual conflict that may distract subjects who could otherwise reason transitively. Second, they argue that the standard procedure of requiring correct verbal explanations for transitive judgments can lead to misclassification; subjects who can reason transitively may often be unable to give lucid explanations of their reasoning processes. Requiring explanations, they argued, can bias the results in favor of subjects with greater verbal facility. Lutkus and Trabasso used a procedure designed to eliminate these problems and to insure that children had learned all necessary preinference information (e.g., that $A > B$). They tested institutionalized retarded adolescents at two MA levels; both groups were apparently unscreened for organic impairment. Overall, the retarded subjects averaged 48 in IQ, 6.0 years in MA, and 15.6 years in CA. These youngsters were compared with nonretarded children from an earlier study (Bryant & Trabasso, 1971) who were thought to be "of comparable MA" (p. 602). Results differed substantially from those of McManis (1969c, 1970) and Gruen (1973). Compared with the nonretarded, retarded subjects did require more trials to learn and remember the necessary preinference information. But once retention of this information was insured, the transitivity performance of the retarded and nonretarded groups did not differ significantly at either MA level.

Number concept development. We have found one comparative study that involved multiple number concepts (Cohn-Jones & Seim, 1978). The investigators apparently did not screen for organic impairment. Retarded and nonretarded samples were both drawn from public schools. The retarded children averaged 72.7 in IQ, 6.8 years in MA, and 10.3 years in CA. In the nonretarded group the mean IQ was 102.2; the mean MA, 6.8 years; and the mean CA, 6.5 years. All subjects were given the Dodwell Number Concept Test (Dodwell, 1960). It is designed to assess abilities in several areas, including conservation of discontinuous quantity, seriation, and serial correspondence. The component measures are combined to yield an overall number concept score and overall explanation scores. Results revealed no significant difference between the two groups on either score.

Studies of Conservation With Etiology Uncontrolled

We now review research in the most frequently studied Piagetian conceptual domain, conservation. Because the studies are so numerous, it is useful to divide them into subgroups; because of the concerns we expressed earlier about the importance of controlling etiology we have grouped the studies according to whether they exercised such control. Cultural–familial retardation accounts for about 75% of all cases of mental retardation (Zigler, 1967); most of the remaining 25% have some form of organic impairment. Consequently, experimenters who do not systematically exclude organically impaired individuals can be expected to have a substantial number of them in any sizeable retarded sample. In dividing studies into two groups, we assigned to the "etiology uncontrolled" group those which reported no effort to eliminate organic cases. Also included in this group are studies where only partial control seems to have been exerted; for example, statements indicating that cases of "gross sensory or motor defect" were eliminated or that the retarded sample was "primarily" familial seem to imply the inclusion of at least some organically impaired children.

First we examine research in the etiology uncontrolled category. The study by DeVries (1970, 1973a, 1973b) included measures of conservation of mass (two tasks), number, length, and continuous quantity (liquid). In addition, constancy of generic identity and sex identity were assessed. DeVries's high-IQ group was significantly superior to her CA-matched average-IQ group on all these measures except one conservation of mass task. Yet the MA-matched average- and low-IQ groups differed significantly on only one of the seven measures. On that measure, sex identity, the retarded group performed significantly better than the nonretarded group.

Kahn's (1976) investigation of moral and cognitive development included measures of conservation of mass and weight. Scores on the two measures were combined to form a composite "cognitive functioning" measure. On this measure, Kahn's nonretarded sample outperformed an MA-matched retarded group, most of whom were organically impaired.

Three studies by McManis included conservation assessments among MA-matched retarded and nonretarded groups. In one, McManis (1969e) sampled institutionalized retarded subjects ranging from 47 to 73 in IQ, from 5.0 to 11.8 years in MA, and from 7.7 to 21.1 years in CA (no means reported). The nonretarded public school sample ranged from 85 to 115 in IQ, 5.3 to 11.9 years in MA, and 5.3 to 10.9 years in CA. Conservation of mass, weight, and volume was assessed. Children were asked to make (a) conservation predictions (e.g., "will there be as much clay in the hotdog as in the ball?", (b) conservation judgments (e.g., "*do* they both have the same amount of clay?"), and (c) explanations of each judgment. These three responses were combined to obtain a score for each type of conservation. Weight conservation scores showed significant effects of MA but did not differentiate MA-matched retarded and nonretarded subjects. There were significant retarded–nonretarded group differences on mass and volume conservation scores, with the nonretarded group superior on mass and the retarded group superior on volume.

McManis (1969c) assessed conservation of weight (clay) and of length (sticks) among MA-matched retarded and nonretarded groups, using a procedure like that in McManis (1969d). He did not report significance tests, but inspection of his data reveals strong MA effects on both measures. However, our own chi-square analysis of McManis's (1969c) frequency data shows that the negligible retarded–nonretarded differences did not approach significance for either type of conservation.

In McManis (1970), MA-matched retarded and nonretarded subjects were given a two-part length conservation task. Part 1 was a relatively standard length conservation procedure. Part 2, patterned after Smedslund (1961), was designed to assess (a) subjects' susceptibility to inducement and extinction of conservation responses and (b) their ability to retain the induced conservation for a 1-week period. Although the MA effect was not tested, the data reported were strongly MA related. However, there was no reliable difference between retarded and nonretarded subjects.

The study by Stevenson et al. (1968), in which children were group-administered a variety of tasks by means of films and test booklets, included groups averaging about 70 and 100 in IQ. The two groups were discribed as comparable in MA, but they were not precisely MA matched. The procedure included a conservation of volume task; subjects were asked to judge the relative amounts of liquid displaced by two lumps of clay with equal mass but dissimilar shapes. Retarded subjects were significantly inferior to the nonretarded.

A conservation study by Goodnow and Bethon (1966) included a group of "dull" and "average" public school boys, matched person-for-person on MA. The dull group had a median IQ of 81, a median MA of 8.6 years, and a median CA of 11.0 years. For the average group the IQ median was 110; the MA median, 8.7 years; and the CA median, 7.8 years. The procedure included measures of conservation of mass, weight, and area. The mass and weight assessments were fairly standard procedures involving lumps of clay. The area measure was more unusual: Subjects compared two fields, one with houses clustered at the side, the other with houses scattered about. In addition, two measures of volume conservation called for children to make judgments about clay balls displacing water. Summing across all conservation measures except mass, the investigators found significant differences between two groups not relevant to our review: high-IQ and average-IQ 11-year-olds. However, summing over all conservation measures, they found no significant difference between the MA-matched groups of average and low IQ. When the measures were analyzed separately, however, three group differences were significant. On the weight measure and the two volume measures, the nonretarded group was significantly superior.

Brown (1973) studied conservation of number and continuous quantity among MA-matched children at three IQ levels. Her bright group averaged 149 in IQ, 6.5 years in MA, and 4.5 years in CA. Her average group had a mean IQ of 102, a mean MA of 6.3 years, and a mean CA of 6.6 years; Her retarded group averaged 74 in IQ, 6.6 years in MA, and 8.4 years in CA. Subjects were drawn from public and university-supported nursery and primary schools in England. On both types of conservation Brown found the bright group significantly inferior to both the average and retarded groups. The retarded and average groups, however, did not differ significantly on either measure. Brown (1973) interpreted this pattern as evidence that "the young child with a high IQ is held back by sheer lack of experience that the additional CA provides" (1973, p. 378). This interpretation is reminiscent of the unconventional difference position advanced by Kohlberg (1968).

Brekke and Williams (1974) administered three conservation of weight measures to retarded and nonretarded children. The nonretarded were from public schools, and the retarded were a mix of public school and institutionalized individuals. Group means were not reported, but IQs ranged from 24 to 86 in the retarded group and from 81 to 160 in the nonretarded group; thus, the groups overlapped in IQ. The retarded group ranged from 3.1 to 12.2 years in MA and from 6.5 to 27.4 years in CA. The nonretarded group ranged from 6.4 to 8.6 years in MA and from 5.5 to 14.9 years in CA. In the 13-step testing sequence, equal-sized balls of clay were taken through various transformations in shape. Subjects were asked to judge the weight equivalence of the clay balls across the successive transformations; they were also asked to explain their judgments (see Furth, 1964, for procedural details). Responses to all conservation questions were combined into one overall conservation of weight score. Scores were higher for nonretarded subjects than for either the institutional or the public school retarded group. This difference was not theoretically meaningful, however; the groups had not been matched for MA, and the nonretarded group had much higher MAs than both retarded groups. To correct for this problem, Brekke and Williams carried out an analysis of covariance (ANCOVA), with MA as covariate. The group differences remained significant. It should be noted, however, that ANCOVAs may pose both conceptual and statistical problems when Piagetian measures are involved (see Kappauf, 1976; Weisz, 1976). One problem is that the assumption of a linear relationship between variate and covariate (Winer, 1971) may be violated if the dependent variable is actually a stagelike phenomenon.

Cardozo and Allen (1975) assessed conservation with the Goldschmid-Bentler (1968) Concept Assessment Conservation Test. They compared institutionalized retarded children with nonretarded public and private school children. The retarded group averaged 63 in IQ, 7.3 years in MA, and 12.2 years in CA. The nonretarded group averaged 107 in IQ, 7.5 years in MA, and 6.9 years in CA. The test used was designed to measure conservation concepts in six different areas: mass, number, weight, continuous and discontinuous quantity, and two-dimensional space. Scores were summed across all six types of conservation. On this composite measure retarded subjects were significantly inferior to the nonretarded contrast group.

Achenbach (1973) assessed one of the earliest manifestations of conservation: children's identity concepts. Identity of color, number, length, and continuous quantity were inferred from children's surprise reactions when these properties were unexpectedly changed. In the test of color identity, for example, five white marbles were inserted into a hole in a box; the box was then turned around, and five black marbles fell out. Children who showed surprise had presumably expected the marbles to have a fixed identity with respect to color. The sample included nonretarded nursery and elementary school children averaging 116 in IQ, 6.3 years in MA, and 5.5 years in CA. The retarded group included both institutional and public school subjects; one fourth suffered from Down's syndrome, but the remainder showed no signs of organic impairment. The Down's syndrome youngsters and the other retarded subjects did not differ in frequency of surprise reactions, so their data were analyzed together. The combined retarded groups averaged 47 in IQ, 6.2 years in MA, and 17.3 years in CA. To fashion a carefully controlled group comparison, Achenbach (1973) formed 49 subject pairs in which a retarded and nonretarded child were matched for MA, sex, race, and order of presentation of the tasks. There were significant differences among the four tasks in the frequency of surprise reactions they evoked; but only one of the four tasks revealed a reliable retarded–nonretarded difference. On the continuous quantity task, retarded subjects showed surprise more frequently than the nonretarded ($p = .058$).[2]

Studies of Conservation, With Etiology Controlled

We now examine conservation studies using retarded samples that were reportedly screened for organic impairment. In one such study, Achenbach (1969) assessed conservation of length, area, and "volume" (actually, continuous and discontinuous quantity in Piagetian terms). He used four different tasks for each property. Each involved an optical illusion that created a discrepancy between the actual and apparent magnitude of a stimulus. In one conservation of area task, for example, the child found a cream colored disk that fit a circular recess in a red piece of plexiglass. Painted arrows pointing to the center of the recess made it look relatively small. Then the child placed the same disk into a second piece of plexiglass; the arrows pointing outward made the recess and the disk seem larger. The child was then asked, (a) which recess looked larger, (b) whether the circle would actually fit back in the original recess, and (c) why it would or would not fit. The retarded subjects averaged 52 in IQ, 6.1 years in MA, and 15.8 years in CA; some were drawn from public schools and others came from state regional institutions. The nonretarded contrast group came from public schools and a private nursery school; they averaged 128 in IQ, 6.8 years in MA, and 5.3 years in CA.

From these two samples 48 retarded–nonretarded subject pairs were formed; pair members were matched for MA, sex, race, and task order. Most of the 12 tasks showed no significant effects of MA or CA. Only one, a discontinuous quantity task, revealed significant IQ-group differences; on it, nine retarded and one nonretarded subject failed to show a conservation response. Achenbach (1969) also scored subjects' explanations of their responses to the 12 items. For 11 of the 12 items, these explanations differentiated significantly between nonretarded subjects of low CA and MA and nonretarded subjects of higher CA and MA, but no significant retarded–nonretarded difference was found on any item.

Keasey and Charles (1967) assessed conservation of mass. Their institutionalized retarded and public school nonretarded subjects were matched person-for-person on MA. IQs were not reported. Retarded subjects ranged from 5.3 to 10.9 years in MA and from 13.0 to 28.7 years in CA. In the nonretarded group the MA range was 5.3–10.9 years, and the CA range was 5.6–10.9 years. Subjects judged whether two identical plasticene balls had the same amount of mass even after transformations in shape. These judgments and their explanations were scored. Both judgments and explanations were significantly related to MA, but neither differentiated significantly between retarded and nonretarded subjects.

Gruen and Vore (1972) used conventional measures to study conservation of number, weight, and continuous quantity. Their sample included retarded children averaging 70 in IQ, 6.9 years in MA, and 10.2 years in CA. An MA-matched nonretarded group averaged 108 in IQ, 7.2 years in MA, and 6.7 years in CA. All subjects came from public schools. The retarded and nonretarded groups were divided into three

[2] The Achenbach (1973) study provides two examples of means by which problems resulting from motivational characteristics of retarded children may be circumvented. The use of brief, minimally verbal techniques for assessing conservation concepts could reasonably be expected to minimize the influence of group differences in motivation to persist (and group differences in verbal facility as well). In addition, Achenbach assessed subjects' IQ and MA scores by means of an "optimizing" procedure. The Stanford-Binet Intelligence Test was administered in a manner calculated to encourage motivation to succeed, with each series of failures followed by simpler problems aimed at producing a success experience, before the examiner proceeded to the next items. Evidence (Zigler, Abelson, & Seitz, 1973; Zigler & Butterfield, 1968) suggests that such a procedure may correct mean underestimates of IQ as high as 10 points in children of lower socioeconomic status (SES). Since the organically intact retarded are almost always from the lower SES, the maximizing procedure may contribute importantly to precision in the formation of experimental groups in any study having a bearing on the similar structure hypothesis.

subgroups, at MAs of 5, 7, and 9 years. The investigators used both conservation-judgment-alone and judgment-plus-explanation as dependent variables. These two measures, assessed across the three types of conservation, yielded numerous significant differences between MA levels and between CA-matched retarded and nonretarded subjects. None of the dependent variables differentiated significantly between retarded subjects and their MA-peers at any of the three MA levels.

Taylor and Achenbach's (1975) study of moral and cognitive development included two conservation tasks. One was a surprise task (from Achenbach, 1973) designed to gauge subjects' color identity concepts. The second was an illusion task (from Achenbach, 1969) designed to gauge conservation of continuous quantity. The continuous quantity task differentiated significantly between MA levels, but the color identity task did not. The MA-matched retarded and nonretarded groups did not differ significantly on either measure.

In Kahn's (1976) sequel to the Taylor–Achenbach study, more traditional measures were used to assess conservation of mass and weight. Scores on both measures were combined to form a cognitive score. Kahn's sample included three MA-matched groups: nonretarded, mildly retarded, and moderately retarded; mean IQs were 101, 66, and 45, respectively. The mildly retarded group was screened for organic impairment; the moderately retarded group included organically impaired individuals. Scores of the nonretarded group were significantly superior to both the moderately and mildly retarded group.[3]

Status of Evidence on the Similar Structure Hypothesis

We have now surveyed Piagetian evidence bearing on the similar structure hypothesis. We examined 30 published experiments involving a total of 104 comparisons between retarded and nonretarded groups with MA controlled. Some 18 different conceptual domains were addressed; these provide good coverage of the preoperational and concrete operational cognitive processes described by Piaget. An uncritical overview of the box-score totals reveals the following: of the 104 group comparisons, only 4 (4%) supported Kohlberg's (1968) unconventional difference position by revealing significantly superior performance among the retarded group; some 25 comparisons (24%) supported the conventional difference position (e.g., Milgram, 1973) by revealing significantly superior performance among the nonretarded group; and 75 comparisons (72%) supported the similar structure hypothesis of the developmental position (Zigler, 1969) by revealing no reliable difference between retarded and nonretarded groups.

Support for the developmental position seems stronger when one critically examines the methodology of these studies. Earlier we stressed the importance of sampling homogeneous groups of retarded subjects by excluding the organically impaired. Of the 104 comparisons, we find 33 that satisfied this condition. Of these 33, only three comparisons (10%) are inconsistent with the similar structure hypothesis. Therefore the simple box-score evidence seems to support the developmental position and its similar structure hypothesis.

Similar Structure Prediction, Power, and the Null Hypothesis

This body of evidence, though, deserves more intense analysis than these simple tallies provide. Several trends may have biased the results. On the one hand, more of the comparisons would probably have been significant if statistical power had been greater. If the studies reviewed had used larger samples, or, perhaps, more precise measurement, some of the nonsignificant group differences might have attained significance. The issue of what is "enough" power is far from resolved; it can be argued from several philosophical and empirical positions (Hays, 1973). Evidence does indicate that psychologists tend to accept relatively low power in their experimental designs (Greenwald, 1975). This tendency may have obscured

[3] This information was provided by James V. Kahn in correspondence with Herman Spitz and with us. We are grateful for this supplement to data reported in Kahn (1976).

trends in the research reviewed, resulting in a spuriously high number of comparisons supporting the similar structure hypothesis.

On the other hand, evidence may have been influenced in the opposite direction by the very nature of the similar structure hypothesis. It is, as we noted earlier, essentially a null hypothesis; that is, it predicts no difference between target groups. Within our discipline, there is a tradition of prejudice against the null hypothesis, a prejudice that may work against the similar structure hypothesis in reviews like this one. We are taught early in our statistics courses that the null hypothesis may be disconfirmed but never confirmed. We learn to regard findings of no difference as failure, and we discover that such findings often remain unpublished. For some time, quantitative social scientists have engaged in a lively debate about the null hypothesis. Statisticians of the Fisher and Yule school regard acceptance of the null hypothesis as inappropriate; statisticians of the Neyman and Pearson school argue that if the null hypothesis can be tested, it also can be accepted as valid (for discussion of the issues involved see, e.g., Binder, 1963, and Grant, 1962).

Greenwald (1975) has argued that the tradition of discrimination against the null may impede scientific progress. There may be inappropriately few publications on problems for which the null hypothesis is a reasonable alternative; a substantial proportion of these will incorrectly reject the hypothesis of no difference. According to this reasoning, findings supporting the similar structure hypothesis may be underrepresented in published reports. This may have been particularly true in the years prior to a clear articulation of the developmental position and its no-difference prediction (Zigler, 1967, 1969). In those years there was no clear theoretical explanation for findings of no difference; this may have undermined the chances that such findings would be published. It is difficult to gauge or correct for the existence of biases for or against the similar structure hypothesis. The existence of such biases, though, does suggest the need for a more critical analysis than the box-score method.

Combining the Evidence

Fortunately, the need for a more rigorous analysis can be met. The task is to use the results of multiple studies to fashion a statistically sound overall test of the similar structure hypothesis. A variety of methods exist (see Rosenthal, 1978). The most precise methods involve pooling exact significance levels and their corresponding Z scores. These methods cannot be used in the absence of very complete statistical data. Such data (e.g., exact significance levels) were not reported in many of the published studies we reviewed; in many cases they were not even available from the authors. Consequently, we selected an alternative approach that allowed us to use all of the studies reviewed while maintaining statistical rigor.[4] We used a counting method that relies on this fact: If a null hypothesis is true, differences between any two groups across independent studies should approximate a normal distribution. This fact means that the proportion of significant differences in either direction found by chance should equal whatever alpha level is selected. A chi-square statistic can thus be used to determine whether the array of differences approximates a normal distribution; it should if the null hypothesis is true. The only requirement for this test is knowledge of the number and direction of significant differences.

All 104 comparisons satisfied this requirement. For our puposes, we chose an alpha level of .05 for both tails of the distribution. The expectation would thus be for 10% of

[4] In an effort to obtain the information needed for pooling of exact significance levels, we mailed requests to the authors, then used follow-up letters and phone calls when necessary. We reached nearly all the authors, but many had discarded or misplaced the data we needed. We also requested information on the direction of all nonsignificant differences between MA-matched retarded and nonretarded groups. Such information would have permitted a four-cell chi-square analysis (i.e., significant and nonsignificant differences favoring retarded and nonretarded groups) that would have allowed us to check for trends in directionality of differences. Even this information was unavailable for almost 20% of the total comparisons reviewed above. We have satisfied ourselves that the analysis we have selected is the most precise one that allows us to include all relevant comparisons.

the findings to be significant (5% on each tail); 90% should not be significant. In other words, if the similar structure hypothesis is true, the comparisons should be distributed as follows: 5% should show the retarded significantly superior to the nonretarded, 90% should show no significant retarded–nonretarded difference, and 5% should show the retarded significantly inferior to the nonretarded. Chi-square statistics were computed comparing the actual to the expected proportions; the statistics were calculated separately for etiology controlled and uncontrolled comparisons.[5] The 104 group comparisons are shown, by category, in Table 1. We classified studies as *etiology controlled* if their authors explicitly stated that organically impaired persons were excluded from the retarded sample. We labeled the remainder, *etiology uncontrolled.* A tabulation of actual and expected frequencies for both categories is shown in Table 2.

The distribution of the etiology uncontrolled comparisons was significantly different from that expected by chance, $\chi^2(2) = 101.53$, $p < .001$; a greater number of comparisons than expected showed differences favoring nonretarded subjects (see Table 1). The distribution of the etiology controlled comparisons was not significantly different from that expected by chance, $\chi^2(2) = 2.76$, $p > .25$. That is, the distribution of comparisons excluding organically impaired retarded persons is not significantly different from that expected if the similar structure hypothesis is true.[6]

Conclusions, Criticisms, and Thoughts on Future Research

Our analysis suggests that when investigators have systematically excluded organically impaired subjects, their findings have generally supported the similar structure hypothesis. When such screening has not been employed, however, findings have supported the conventional difference position. The similar structure hypothesis has been clearly articulated as applicable only to nonorganically impaired children (cf. Weisz & Zigler, 1979; Zigler 1967, 1969); thus, the evidence must be regarded as substantially in support of the hypothesis. An earlier review of Piagetian research on the other major facet of the developmental position (i.e., the similar sequence hypothesis) revealed that it too was supported by the findings (see Weisz & Zigler, 1979). In summary then, a substantial body of Piagetian data supports both major tenets of the developmental position.

If in the future the developmental position is supported by other bodies of research not reviewed here or by Weisz and Zigler (1979), significant implications would follow for the ways we construe and diagnose mental retardation. One implication would be that mentally retarded persons who do not suffer from specific organic deficits should not be construed as fundamentally abnormal. Instead, they might best be described as man-

[5] The expected cell frequencies for these analyses are lower than those frequently recommended as a limit on the applicability of the chi-square statistic (e.g., Hays, 1973). Recent research (Camilli & Hopkins, 1976; Roscoe & Byars, 1971), however, has shown that the chi-square statistic is very robust in the face of violations of these limits. Indeed, the research suggests that expected cell frequencies as low as 1 or 2 produce valid statistics and significance levels as long as total sample size is 20 or more. Our analyses fulfill these conditions. We felt that the analysis was further justified because recent research (Cooper & Rosenthal, 1980) suggested that literature reviews such as ours, using statistical procedures for combining studies, provide more rigorous conclusions than those employing subjective judgments.

[6] Spitz (in press), in a critique of the developmental position, has argued that four of the studies included in this table are questionable as regards the diagnostic categories in which we have placed them (i.e., whether or not they include organically impaired retarded persons). Although we agree with Spitz that the sample descriptions given in some of the studies do not permit complete certainty, we have adopted what seems to us a reasonable policy of classifying studies as having excluded organically impaired persons if their authors stated explicitly that this procedure was followed. Nonetheless, as a check on whether the experiments questioned by Spitz might have distorted the pattern of results, we recalculated the chi-square analyses depicted in Table 2, excluding the four studies regarded by Spitz as doubtful (Achenbach, 1969; Cardozo & Allen, 1975; Goodnow & Bethon, 1966; Stevenson et al., 1968). The pattern of results was similar to that shown in Table 2. The table for comparisons in which organic subjects were not systematically excluded was again significant, $\chi^2(2) = 57.35$, $p < .001$; the table for comparisons in which organic subjects were excluded was again nonsignificant, $\chi^2(2) = 1.91$, $p > .25$.

Table 1
Directional Comparisons of MA-Matched Retarded and Nonretarded Persons on Various Piagetian Tasks

Study	Direction of difference		
	MR > NMR	MR ≅ NMR	MR < NMR
		Etiology controlled	
Achenbach (1969)		Conservation of length × 4, area × 4, continuous quantity × 2, discontinuous quantity	Conservation of discontinuous quantity[a]
Gruen (1973)		Transitivity (no memory aid)	Transitivity (memory aid)
Gruen & Vore (1972)		Conservation of number × 3, weight × 3, continuous quantity × 3	
Kahn (1976)		Moral judgement	Conservation of mass and weight (combined)
Keasey & Charles (1967)		Conservation of mass	
Taylor & Achenbach (1975)		Conservation of continuous quantity Color identity, relative thinking, role taking × 2, seriation, moral judgment	
		Etiology uncontrolled	
Achenbach (1973)	Conservation of continuous quantity	Conservation of number, color, length	
Brekke & Williams (1974)			Conservation of weight
Brown (1973)		Conservation of number, continuous quantity	
Cardozo & Allen (1975)		Conservation of mass, weight, two-dimensional space, continuous and discontinuous quantity (combined)	
Cohn-Jones & Seim (1978)		Multiple number concepts, number concept judgments	
DeVries (1970, 1973a, 1973b)	Sex identity	Conservation of mass × 2, number, length, continuous quantity Generic identity, classification, class inclusion, transitivity, relative thinking × 2, dream concepts, magic	Role taking
Goodnow & Bethon (1966)		Conservation of mass, area	Conservation of weight, volume × 2
Granich (1940)[b]		Animism, artificialism, realism	

Table 1 (*Continued*)

Study	Direction of difference		
	MR > NMR	MR ≅ NMR	MR < NMR
Kahn (1976)			Moral judgment Conservation of mass and weight (combined)
Lane & Kinder (1939)[c]		Relative thinking × 2	
Lutkus & Trabasso (1974)		Transitivity × 2	
McManis (1968)		Additive composition	
McManis (1969a)		Gross quantity	Intensive and extensive quantity
McManis (1969c)		Conservation of weight, length	Transitivity × 2
McManis (1969e)	Conservation of volume	Conservation of weight	Conservation of mass
McManis (1969f)[b]			Relative thinking × 2
McManis (1970)		Conservation of length	Transitivity, seriation
Prothro (1943)		Animism	Relative thinking, classification, causality
Rubin & Orr (1974)		Perspective taking × 2	
Russell, Dennis, & Ash (1940)	Animism		
Smeets (1973, 1974)		Life attributions × 2, life trait attributions × 2	
Smith (1977a)		Logical contradiction	
Smith (1977b)		Perceptual decentering	
Stevenson, Hale, Klein, & Miller (1968)[d]			Probability × 2 Conservation of volume

Note. MA = mental age; MR = mentally retarded subjects; NMR = nonretarded subjects; the symbol ≅ indicates that there was no reliable group difference.
[a] This result was only true for initial conservation *response*. Conservation *explanation* scores on this task did not differentiate significantly between retarded and nonretarded groups. [b] The directional designation of each comparison from this study in the table relies on the judgments of the respective authors because no significance tests were reported. [c] Lane and Kinder actually compared two groups of retarded persons, matched for MA but differing significantly in IQ. [d] In this study, MA matching is only approximate and MA differences between retarded and nonretarded groups were not actually tested for significance.

ifesting a developmental lag. According to this view, they pass through a normal sequence of cognitive developmental stages. Their rate of progress through that sequence is relatively slow, and their upper limit of cognitive development is relatively low. However, when a nonorganically impaired retarded person and a nonretarded person are at the same level of development, their reasoning derives from similar underlying cognitive structure.

If this perspective is valid, then the diagnosis of mental retardation might best be conditioned on the magnitude of developmental lag. Prognostic expectations would depend on rate of cognitive development. According to this reasoning, the psychometric MA may be used to make rough estimates of relative cognitive deficiency at a given point in time. The psychometric IQ has been described elsewhere as an index of lifetime learning rate (Harter, 1965; Weir,

Table 2
Directional Comparisons of MA-Matched Retarded and Nonretarded Persons Summed Across Various Piagetian Tasks

	Direction of difference		
Group	MR > NMR	MR ≅ NMR	MR < NMR
Etiology controlled[a]			
Actual total	0	30	3
Expected total	1.65	29.7	1.65
Etiology uncontrolled[b]			
Actual total	4	45	22
Expected total	3.55	63.9	3.55

Note. MA = mental age; MR = mentally retarded subjects; NMR = nonretarded subjects; the symbol ≅ indicates that there was no reliable group difference.
[a] χ^2 (2) = 2.76, $p > .25$. [b] χ^2 (2) = 101.53, $p < .001$.

1967); it may provide a useful approximation of the relative pace of development.

This view of mental retardation might also carry implications for mainstreaming programs. Such programs often yield classes that are relatively homogeneous with respect to physical development but heterogeneous in cognitive maturity. Such an arrangement may pose problems for both teachers and pupils (Zigler & Meunchow, 1979). The findings reviewed here and in Weisz and Zigler (1979) raise a question as to whether MA should be a basis for mainstreaming in some cases. One example may be classes where a low teacher–pupil ratio precludes one-to-one instruction and where the subject matter is cognitively demanding for the majority of the class. In such classes retarded children similar to the majority in CA are not likely to gain much intellectually; they also may suffer some loss of self-esteem. In cases like this, mainstreaming retarded children with nonretarded children of similar MA would offer certain advantages. There would be disadvantages as well. For example, retarded and nonretarded children would continue to differ in significant ways that would need to be taken into account in educational planning; differences would certainly exist in physical development, and perhaps also in the ability to utilize cognitive structures in learning situations. However, mainstreaming based on MA would at least allow teachers to employ similar ability expectations for their retarded and their nonretarded students[7]; it might also prevent retarded children from being repeatedly placed at a competitive disadvantage.

All these speculations, however, are based on the assumption that the developmental position is valid. Such an assumption is premature. Research bearing on the developmental position needs to be improved in a number of ways. Investigators have paid little attention to personality factors that can undermine performance in retarded children (see e.g., Weisz, in press-b; Zigler, 1971). Few of the studies reviewed here included any explicit effort to maximize motivation or self-confidence prior to the Piagetian tasks. Moreover, a substantial number of the retarded subjects resided in institutions; a number of these were adults, some middle-aged. Research has demonstrated significant depressing effects of institutionalization on learning activity (see Denny, 1964; Harter, 1967), quality of language behavior (see Lyle, 1959), and level of abstraction in vocabulary (see Badt, 1958). In addition to these environmental effects, one can imagine

[7] Of course this position assumes that teachers would employ similar ability expectations. Recent evidence suggests that this may not be the case. Adults attribute the failure of a mentally retarded child to low ability, but they attribute the failure of an MA-matched nonretarded child to low effort (Weisz, in press-a). This pattern may represent an attributional bias; adults may overgeneralize their perception of the retarded child as "low in ability." Or it may be that most adults hold an implicit difference position like that described in this article. In either case, adults, particularly potential parents and teachers of the retarded, may benefit from evidence such as that presented in the present review.

motivational problems arising in the retarded adults who were asked to work with modeling clay or to engage in other Piagetian tasks designed for young children.

Much of the evidence can be faulted for imprecise MA matching and for a failure to control for organic impairment. Studies that evidenced such control yielded a markedly different pattern of findings than studies that did not. This difference in pattern might be seen as support for Zigler's (1967) "two group" approach. That approach takes the view that organically impaired retarded persons are cognitively different from those not suffering from specific organic deficits. Following this line of reasoning one could argue that much of the support found for the conventional difference position is an artifact of organic impairment in the retarded samples. The data reviewed do not clearly warrant such a conclusion. The mean IQs in the samples that included organically impaired subjects were generally lower than the mean IQs of the screened samples. Thus, as Spitz (in press) has noted, IQ level and organicity are generally confounded across the studies reviewed. To generate more definitive evidence on the two-group approach in the future, investigators will need to accomplish a difficult task: disentangling organicity and IQ level by structuring designs in which the two factors are orthogonal.

Perhaps the most important task for the future is to survey non-Piagetian research. Evidence from the Genevan tradition has a vigorous cognitive developmental emphasis. It is thus the logical starting point for reviews on the status of the developmental versus difference controversy. But the controversy also addresses processes of learning and reasoning that lie outside the range of Piagetian theory. Some of these processes (discussed in Zigler, 1967, 1969) have stimulated substantial research by now. They warrant critical analyses of the sort undertaken here. Such analyses are needed if we are to consolidate the gains in knowledge stimulated by the developmental versus difference debate. For the time being, any answer to the question raised at the outset of this article must be carefully qualified. When retarded persons suffering from organic impairment are excluded, retarded and nonretarded persons of the same developmental level do seem to use the same cognitive processes, at least in the cognitive domains addressed by Piaget. The returns from other domains are not yet in.

Reference Note

1. Kuhn, D. *The development of role-taking ability.* Unpublished manuscript, Columbia University, 1972.

References

Achenbach, T. M. Conservation of illusion-distorted identity: Its relation to MA and CA in normals and retardates. *Child Development*, 1969, *40*, 663–679.

Achenbach. T. M. Comparison of Stanford-Binet performance of nonretarded and retarded persons matched for MA and sex. *American Journal of Mental Deficiency*, 1970, *74*, 488–494.

Achenbach, T. M. Stanford-Binet Short-Form performance of retarded and nonretarded persons matched for MA. *American Journal of Mental Deficiency*, 1971, *76*, 30–32.

Achenbach, T. M. Surprise and memory as indices of concrete operational development. *Psychological Reports*, 1973 *33*, 47–57.

Badt, M. I. Levels of abstraction in vocabulary definitions of mentally retarded school children. *American Journal of Mental Deficiency*, 1958, *63*, 241–246.

Balla, D., Styfco, S. J., & Zigler, E. Use of the opposition concept and outerdirectedness in intellectually average, familial retarded, and organically retarded children. *American Journal of Mental Deficiency*, 1971, *75*, 663–680.

Balla, D., & Zigler, E. Discrimination and switching learning in normal, familial retarded, and organic retarded children. *Journal of Abnormal and Social Psychology*, 1964, *61*, 664–669.

Binder, A. Further consideration on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, 1963, *70*, 107–115.

Blount, W. R. Concept usage research with the mentally retarded. *Psychological Bulletin*, 1968, *69*, 281–294.

Brekke, B., & Williams, J. D. Conservation of weight with the mentally retarded. *Journal of Genetic Psychology*, 1974, *125*, 225–231.

Brown, A. L. Conservation of number and continuous quantity in normal, bright, and retarded children. *Child Development*, 1973, *44*, 376–379.

Bryant, P. E., & Trabasso, T. Transitive inferences and memory in young children. *Nature*, 1971, *232*, 456–458.

Camilli, G., & Hopkins, K. D. Applicability of chi-square to 2 × 2 contingency tables with small expected cell frequencies. *Psychological Bulletin*, 1976, *85*, 163–167.

Cardozo, C. W., & Allen, R. M. Contribution of visual perceptual maturation to the ability to conserve. *American Journal of Mental Deficiency*, 1975, *79*, 701–704.

Cohn-Jones, L., & Seim, R. Perceptual and intellectual factors affecting number concept development in retarded and nonretarded children. *American Journal of Mental Deficiency*, 1978, *83*, 9–15.

Cole, M., & Bruner, J. Cultural differences and inferences about psychological processes. *American Psychologist*, 1971, *26*, 867–876.

Cooper, H. M., & Rosenthal, R. Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 1980 *87*, 442–449.

Cromwell, R. L. A social learning approach to mental retardation. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.

Cruikshank, W. M. *The brain-injured child in home, school, and community*. Syracuse, N.Y.: Syracuse University Press, 1967.

Denny, M. R. Research in learning and performance. In J. A. Stevens & R. Heber (Eds.), *Mental retardation: A review of research*. Chicago: University of Chicago Press, 1964.

DeVries, R. The development of role-taking as reflected by behavior of bright, average, and retarded children in a social guessing game. *Child Development*, 1970, *41*, 759–770.

DeVries, R. *Performance on Piaget-type tasks of high-IQ, average-IQ, and low-IQ children*. Chicago: University of Illinois at Chicago Circle, 1973. (ERIC Document Reproduction Service No. ED086 374/PS007 129) (a)

DeVries, R. *The two intelligences of bright, average, and retarded children*. Chicago: University of Illinois at Chicago Circle, 1973. (ERIC Document Reproduction Service No. ED 079/102/SE 016 419) (b)

Dodwell, P. Children's understanding of number and related concepts. *Canadian Journal of Psychology*, 1960, *14*, 191–205.

Elkind, D., Koegler, R., & Go, E. Studies in perceptual development: II. Part-whole perception. *Child Development*, 1964, *35*, 81–90.

Elkind, D., Koegler, R. R., Go, E., & Van Doorninck, W. Effects of perceptual training on unmatched samples of brain-injured and familial retarded children. *Journal of Abnormal Psychology*, 1965, *70*, 107–110.

Ellis, N. R. A behavioral research strategy in mental retardation: Defense and critique. *American Journal of Mental Deficiency*, 1969, *73*, 557–566.

Flavell, J. H., Botkin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. *The development of role-taking and communication skills in children*. New York: Wiley, 1968.

Furth, H. G. Conservation of Weight in deaf and hearing children. *Child Development*, 1964, *35*, 143–150.

Gibson, B. J. *An attributional analysis of performance outcomes and alleviation of learned helplessness on motor performance tasks: A comparative study of educable mentally retarded and nonretarded boys*. Unpublished doctoral dissertation, University of Alberta, 1980.

Goldschmid, M., & Bentler, P. M. The dimension and measurement of conservation. *Child Development*, 1968, *39*, 787–802.

Goodnow, J. J., & Bethon, G. Piaget's tasks: The effects of schooling and intelligence. *Child Development*, 1966, *37*, 573–582.

Granich, L. A qualitative analysis of concepts in mentally deficient schoolboys. *Archives of Psychology*, 1940, *251*, 1–47.

Grant, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, 1962, *69*, 54–61.

Greenwald, A. G. Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 1975, *82*, 1–20.

Gruen, G. E. Memory, IQ, and transitive inference in normals and retardates. *Developmental Psychology*, 1973, *9*, 436.

Gruen, G. E., & Vore, D. A. Development of conservation in normal and retarded children. *Developmental Psychology*, 1972, *6*, 146–157.

Harter, S. Discrimination learning set in children as a function of IQ and MA. *Journal of Experimental Child Psychology*, 1965, *2*, 31–43.

Harter, S. Mental age, IQ, and motivational factors in the discrimination learning set performance of normal and retarded children. *Journal of Experimental Child Psychology*, 1967, *5*, 123–141.

Harter, S., Brown, L., & Zigler, E. The discrimination learning of normal and retarded children as a function of penalty conditions and etiology of retarded. *Child Development*, 1971, *42*, 517–536.

Hays, W. L. *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston, 1973.

Heber, R. T. Personality. In H. A. Stevens & R. T. Heber (Eds.), *Mental retardation: A review of research*. Chicago: University of Chicago Press, 1964.

Inhelder, B. *The diagnosis of reasoning in the mentally retarded*. New York: Day, 1968. (Originally published, 1943).

Kahn, J. V. Moral and cognitive development in moderately retarded, mildly retarded, and nonretarded individuals. *American Journal of Mental Deficiency*, 1976, *81*, 209–214.

Kamin, L. J. *The science and politics of IQ*. Potomac, Md.: Erlbaum, 1975.

Kappauf, W. E. Critique of the use of covariance adjustments for CA and MA in comparative studies of retarded and nonretarded persons. *American Journal of Mental Deficiency*, 1976, *81*, 240–247.

Keasey, C. T., & Charles, D. C. Conservation of substance in normal and retarded children. *Journal of Genetic Psychology*, 1967, *111*, 271–279.

Kohlberg, L. Early education: A cognitive-developmental view. *Child Development*, 1968, *39*, 1013–1062.

Kohlberg, L. Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally, 1969.

Labov, W. The logical non-standard English. In F. Williams (Ed.), *Language and poverty*. Chicago: Markham Press, 1970.

Lane, E. B., & Kinder, E. F. Relativism in the thinking of subnormal subjects as measured by certain of Piaget's tests. *Journal of Genetic Psychology*, 1939, *54*, 107–118.

Luria, A. R. Psychological studies of mental deficiency in the Soviet Union. In N. R. Ellie (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.

Lutkus, A., & Trabasso, T. Transitive inferences by preoperational retarded adolescents. *American Journal of Mental Deficiency*, 1974 *78*, 599–606.

Lyle, L. The effect of an institutional environment upon the verbal development of imbecile children. I. Verbal intelligence. *Journal of Mental Deficiency Research*, 1959, *3*, 122–128.

Mahaney, E. J., & Stephens, B. Two-year gains in moral judgment by retarded and nonretarded persons. *American Journal of Mental Deficiency*, 1974, *79*, 134–141.

McManis, D. L. Additive composition by normal and retarded children of matched MA. *Psychonomic Science*, 1968, *11*, 207.

McManis, D. L. Comparisons of gross, intensive, and extensive quantities by normals and retardates. *Child Development*, 1969, *40*, 237–244 (a)

McManis, D. L. Comparisons of gross, intensive, and extensive quantities by retardates. *Journal of Genetic Psychology*, 1969, *115*, 229–236. (b)

McManis, D. L. Conservation and transitivity of weight and length by normals and retardates. *Developmental Psychology*, 1969, *1*, 373–382. (c)

McManis, D. L. Conservation of identity and equivalence of quantity by retardates. *Journal of Genetic Psychology*, 1969, *115*, 63–69. (d)

McManis, D. L. Conservation of mass, weight, and volume by normal and retarded children. *American Journal of Mental Deficiency*, 1969, *73*, 762–767. (e)

McManis, D. L. Relative thinking by normals and retardates. *Developmental Psychology*, 1969, *1*, 69. (f)

McManis, D. Conservation, seriation, and transitivity performance by retarded and average individuals. *American Journal of Mental Deficiency*, 1970, *74*, 784–791.

Milgram, N. A. The rationale and irrational in Zigler's motivational approach to mental retardation. *American Journal of Mental Deficiency*, 1969, *73*, 527–532.

Milgram, N. A. Cognition and language in mental retardation: Distinctions and implications. In D. K. Routh (Ed.), *The experimental psychology of mental retardation*. Chicago: Adline, 1973.

Piaget, J. *Judgment and reasoning in the child*. New York: Harcourt, Brace, 1928.

Piaget, J. *The child's conception of the world*. New York: Harcourt, Brace, 1929.

Piaget, J. *The child's conception of physical causality*. London: Kegan Paul, Trench, & Trubner, 1930.

Piaget, J. *The child's conception of number*. New York: Humanities Press, 1952.

Piaget, J. Cognitive development in children: The Piaget papers. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget rediscovered: A report of the Conference on Cognitive Studies and Curriculum Development*. Ithaca, N.Y.: Cornell University, School of Education, 1964.

Piaget, J. Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (3rd ed.). New York: Wiley, 1970.

Prothro, E. T. Egocentricity and abstraction in children and in adult aments. *American Journal of Psychology*, 1943, *56*, 66–77.

Reitan, R. M. Psychological assessment of deficits associated with brain lesions in subjects. In J. L. Khanna (Ed.), *Brain damage and mental retardation: A psychological evaluation*. Springfield, Ill.: Charles C Thomas, 1973.

Roscoe, J. T., & Byars, J. A. An investigation of the restraints with respect to sample size commonly imposed on the use of the chi-square statistic. *Journal of the American Statistical Association*, 1971, *66*, 755–759.

Rosenthal, R. Combining results of independent studies. *Psychological Bulletin*, 1978, *85*, 185–193.

Rubin, K. H., & Orr, P. R. Spatial egocentrism in nonretarded and retarded children. *American Journal of Mental Deficiency*, 1974, *79*, 95–97.

Russell, R. W., Dennis, W., & Ash, F. E. Studies in animism: III Animism in feebleminded subjects. *Journal of Genetic Psychology*, 1940, *57*, 57–63.

Sarason, S. B., & Doris, J. *Psychological problems in mental deficiency*. New York: Harper & Row, 1969.

Smedslund, J. The acquisition of conservation of substance and weight in children: VI. Practice on continuous versus discontinuous material in conflict situations without external reinforcement. *Scandinavian Journal of Psychology*, 1961, *2*, 203–210.

Smeets P. M. The animism controversy revisited: A probability analysis. *Journal of Genetic Psychology*, 1973, *123*, 219–225.

Smeets, P. M. The influence of MA and CA on the attribution of life and life traits to animate and inanimate objects. *Journal of Genetic Psychology*, 1974, *124*, 17–27.

Smith, J. D. EMR and nonretarded children's reactions to contradiction. *American Journal of Mental Deficiency*, 1977, *82*, 94–94. (a)

Smith, J. D. Perceptual decentering in EMR and nonretarded children. *American Journal of Mental Deficiency*, 1977, *81*, 499–501. (b)

Spitz, H. H. Critique of the developmental position. In E. Zigler & D. Balla (Eds.), *Mental retardation: The developmental–difference controversy*. Hillsdale, N.J.: Lawrence Erlbaum, in press.

Stephens, B. *The development of reasoning, moral judgment, and moral conduct in retardates and normals: Phase II*. Philadelphia, Pa.: Temple University, 1972.

Stephens, B., & McLaughlin, J. A. Two-year gains in reasoning by retarded and nonretarded persons. *American Journal of Mental Deficiency*, 1974, *79*, 116–126.

Stephens, B., et al. Symposium: Developmental gains in the reasoning, moral judgment, and moral conduct of retarded and nonretarded persons. *American Journal of Mental Deficiency*, 1974, *79*, 113–161.

Stevenson, H. W. Social reinforcement of children's behavior. In L. P. Lipsitt & C. Spiker (Eds.), *Advances in child development* (Vol. 2) New York: Academic Press, 1965.

Stevenson, J. W., Hale, G. A., Klein, R., & Miller, L. Interrelation and correlations in children's learning and problem solving. *Monographs of the Society for*

Research in Child Development, 1968, *33*(7, Serial No. 123).

Taylor, J. J., & Achenbach, T. M. Moral and cognitive development in retarded and nonretarded children. *American Journal of Mental Deficiency*, 1975, *80*, 43–50.

Tulkin, S. R., & Konner, M. J. Alternative conceptions of intellectual functioning. *Human Development*, 1973, *16*, 33–52.

Weir, M. Mental retardation. *Science*, 1967, *157*, 576–577.

Weisz, J. R. Studying cognitive development in retarded and nonretarded groups: The role of theory. *American Journal of Mental Deficiency*, 1976, *81*, 235–239.

Weisz, J. R. A follow-up developmental study of hypothesis behavior among mentally retarded and nonretarded children. *Journal of Experimental Child Psychology*, 1977, *34*, 108–122.

Weisz, J. R. Perceived control and learned helplessness among mentally retarded and nonretarded children: A developmental analysis. *Developmental Psychology*, 1979, *15*, 311–319.

Weisz, J. R. Learned helplessness in black and white children identified by their schools as retarded and nonretarded: Performance deterioration in response to failure. *Developmental Psychology*, 1981, *4*, 385–390.

Weisz, J. R. Effects of the "mentally retarded" label on adults' attributions for child failure. *Journal of Abnormal Psychology*, in press. (a)

Weisz, J. R. Learned helplessness and the retarded child. In E. Zigler & D. Balla (Eds.), *Mental retardation: The developmental-difference controversy*. Hillsdale, N.J.: Lawrence Erlbaum, in press. (b)

Weisz, J. R., & Achenbach, T. M. Effects of IQ and mental age on hypothesis behavior in normal and retarded children. *Developmental Psychology*, 1975, *11*, 304–310.

Weisz, J. R., Yeates, K. O., & Zigler, E. Piagetian evidence and the developmental–difference controversy. In E. Zigler & D. Balla (Eds.), *Mental retardation: The developmental-difference controversy*. Hillsdale, N.J.: Lawrence Erlbaum, in press.

Weisz, J. R., & Zigler, E. Cognitive development in retarded and nonretarded persons: Piagetian tests of the similar sequence hypothesis. *Psychological Bulletin*, 1979, *86*, 831–851.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Zigler, E. Familial mental retardation: A continuing dilemma. *Science*, 1967, *155*, 292–298.

Zigler, E. Developmental versus difference theories of mental retardation and the problem of motivation. *American Journal of Mental Deficiency*, 1969, *73*, 536–556.

Zigler, E. The retarded child as a whole person. In H. E. Adams & W. K. Boardman (Eds.), *Advances in experimental clinical psychology*. New York: Pergamon Press, 1971.

Zigler, E., Abelson, W., & Seitz, V. Motivational factors in the performance of economically disadvantaged children on the Peabody Picture Vocabulary Test. *Child Development*, 1973, *44*, 294–303.

Zigler, E., & Butterfield, E. C. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. *Child Development*, 1968, *39*, 1–14.

Zigler, E., & Harter, S. Socialization of the mentally retarded. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally, 1969.

Zigler, E., & Meunchow, S. Mainstreaming: The proof is in the implementation. *American Psychologist*, 1979, *34*, 993–996.

Received February 2, 1981 ■