JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY **24**, 108–122 (1977)

A Follow-Up Developmental Study of Hypothesis Behavior Among Mentally Retarded and Nonretarded Children

JOHN R. WEISZ

Yale University

To clarify the roles of IQ and mental age (MA) in hypothesis behavior, MA-matched subjects at three levels of IQ (70, 100, and 130) and three levels of MA (5½, 7½, and 9½ years) received blank-trial discrimination learning problems using procedures designed to discourage position-oriented responding. With position responding discouraged, earlier findings were contradicted in that no hypothesis measure showed a main effect of IQ. This suggests that previously reported IQ group differences in hypothesis behavior may not reflect cognitive deficits inherently linked to low IQ, but instead may reflect the influence of specific methodological factors. The finding and interpretation are consistent with Zigler's (*American Journal of Mental Deficiency*, 1969, **73**, 536–556) "developmental" theory of retardation and inconsistent with the general "difference" position. In additional findings, the predictions that subjects at all three MA levels would use hypotheses, and that retarded children from special-education classes would use hypotheses more often than retarded children "mainstreamed" in classes for the nonretarded were confirmed.

In cognitive research with normal and atypical populations, an important issue concerns the effects of rate and level of development on intellectual activity. A major theoretical controversy revolves around the question of whether individuals who differ in overall rate of development are dissimilar intellectually at that point in time when they have attained the same level of development. On the one hand, Zigler (1969) has advanced a "developmen-

This study is based on a Ph.D. dissertation presented to the Psychology Department of Yale University. It was supported by a grant-in-aid-of-research from Sigma Xi, the Scientific Research Society of North America; a research grant from the Department of Psychology at Yale; and National Institute of Child Health and Human Development Grant HD03008. The author was supported through National Institute of Mental Health Predoctoral Research Fellowship 1 F01 MH54552-01.

I am most grateful to committee members Thomas Achenbach, Edward Zigler, Irvin Child, Jerome Singer, and Louis Heifetz, as well as Fred Rothbaum, for generous doses of sage advice. Special thanks are also due to Hildy Feen, John Hill, and the administration, teachers, and pupils of the Ansonia, Connecticut, and West Haven, Connecticut, school systems.

Requests for reprints or an extended description of the procedure should be addressed to John R. Weisz, Department of Human Development and Family Studies, Cornell University, Ithaca, N.Y. 14853.

ISSN 0022-0965

tal" position, intended to apply whenever organic impairment is not involved. He maintains that persons differing in developmental rate (operationally defined as IQ) but equated for developmental level (operationally defined as mental age, or MA) will not differ in the formal cognitive processes they employ in learning and reasoning. Numerous other investigators hold what Zigler had labeled the "difference" position—the view that individuals of differing IQ but similar MA will manifest many cognitive differences inherently related to IQ. Most difference positions generate the prediction that subjects of low IQ will be cognitively inferior to subjects of similar MA but higher IQ on many cognitive tasks. For example, Zeaman (1973) maintains that retarded children are inferior to MA-matched nonretarded children in attention and retention (processes critical to hypothesis testing), while Zigler argues that no such differences exist. An extended discussion of the developmental vs difference controversy is available elsewhere (Weisz & Zigler, Note 2).

The developmental vs difference controversy has generated an extensive but inconclusive literature on performance of MA-matched groups differing in IQ. (See Denny, 1964; Heal & Johnson, 1970; Weisz, 1976; Weisz & Zigler, Note 2; Zeaman & House, 1967). To explain the inconclusiveness of the literature, Harter (1965) and Weir (1967) have argued that IQ reflects lifetime learning rate and will also predict rate of learning in an experiment if the experimental task is sufficiently complex. To test the Harter–Weir prediction that complex tasks are more likely than simple tasks to show inferior learning ability in the retarded, Weisz and Achenbach (1975) administered relatively simple, two-dimensional discrimination learning tasks and more complex, four-dimensional tasks to groups of MA-matched retarded and intellectually average subjects. A "blank-trials" procedure (cf. Levine, 1966) was used, by which feedback was withheld on selected blocks of trials. The term "hypothesis" was operationally defined as a consistent selection of one stimulus property (e.g., the color blue) across a series of nonfeedback trials. Thus, without requiring subjects to state their hypotheses, the procedure indicated whether (a) subjects used hypotheses, and (b) the hypotheses changed following feedback.

Two major findings of the study were consistent with the Harter–Weir prediction. On complex problems retarded subjects proved inferior in the number of hypothesis patterns used and in the tendency to repeat a hypothesis after feedback indicating that a response was correct; on the simpler problems there was no group difference on any hypothesis measure.

The Weisz–Achenbach findings could be pivotal since they not only support the Harter–Weir hypothesis, but also suggest a process-related explanation for a pattern which characterizes much of the literature on the developmental vs difference controversy, i.e., studies using relatively simple tasks have shown IQ effects less often than studies using more complex tasks. It now seems that this interaction of IQ level with task complexity may be accounted for by differences in hypothesis behavior during learning—differences which depend upon task complexity. One interpretation of the interaction reported by Weisz and Achenbach and the similar pattern often seen in related literature is that group differences on the simpler tasks in each instance are attenuated by "floor effects." What this interpretation leaves unexplained, of course, is the origin of IQ group differences on relatively complex tasks. An important first step in this regard, with respect to the Weisz–Achenbach findings, is to determine whether such IQ group differences in hypothesis behavior reflect cognitive deficiencies linked to low IQ (a *difference* interpretation), or instead reflect the effects of specific experimental procedures (a *developmental* interpretation; cf. Zigler, 1969); there is reason to suspect the latter.

EFFECTS OF EXPERIMENTAL METHODOLOGY

The blank-trials method employed by Weisz and Achenbach (1975), like that of several earlier investigators (e.g., Levine, 1966; Eimas, 1969), provided that the position of stimuli on each stimulus card (i.e., left or right side of the card) be described to subjects as a possible basis for the solution of the discrimination problems. Thus, a subject's choice of, say, the left stimulus on each trial of a blank-trial block was taken to indicate the use of a legitimate hypothesis. In addition, other position-oriented response sequences such as alternation and double alternation often coincided with legitimate hypotheses.

Gholson, Levine, and Phillips (1972) have argued that scoring such position responses as hypotheses can lead to misinterpretation of results. These investigators, employing a procedure in which position perseveration, alternation, and double alternation were disallowed, found group means for several measures of hypothesis behavior that differed rather widely from those of comparable age groups in the Weisz–Achenbach study. This raises a question as to the validity of the retarded–nonretarded group difference reported by Weisz and Achenbach, but as yet it is unclear whether discrepancies between findings of these two studies resulted from a failure by Weisz and Achenbach to disallow position responding or a failure by Gholson et al. to produce a highly motivating test situation (see discussion by Weisz & Achenbach, 1975). Thus, in order to evaluate properly the Weisz–Achenbach findings, it was necessary both (a) to disallow position responses, as was done by Gholson et al., and b) to employ highly motivating experimental conditions as was done by Weisz and Achenbach.

It should also be noted that Gholson et al. disallowed position-oriented responses only by failing to mention stimulus position when the possible solutions were outlined for subjects. The failure to exclude this prepotent stimulus dimension in a more direct manner raises an important question, since a number of investigators have, unlike Gholson et al., interpreted spontaneous, systematic position responding as evidence that strategies (cf. Gruen & Zigler, 1968) and hypotheses (cf. Rieber, 1969) *are* being used. Since the interpretation of position-oriented responding is a matter of controversy at the present time, and since a vast number of nonposition dimensions can easily be employed in the blank-trials procedure, it was deemed appropriate to actively direct children's attention to stimulus dimensions other than the position of stimuli on the stimulus card.

So, the present study employed two procedures for assessing hypothesis behavior. All subjects received four-dimensional discrimination problems similar to those of Weisz and Achenbach. In pretraining, position was omitted from the list of relevant dimensions, and subjects were informed explicitly that only the dimensions listed would be relevant. Half the subjects used stationary stimulus cards with each card containing two stimuli. The remaining subjects used the same cards rotating on a turntable so that relative stimulus positions were constantly in flux and position responding was actively (though nonverbally) discouraged.

The three MA levels employed by Weisz and Achenbach (5½, 7½, and 9½ years) were completely crossed with three IQ levels (70, 100, and 130) instead of the two levels used in the earlier study. The range of IQ was expanded to increase the sensitivity of the various measures to IQ effects and to provide data on children of superior IQ, a group which has received relatively little attention in developmental research. The three MA levels were included (a) so that findings could be compared directly with the findings of Weisz and Achenbach, and (b) to assess the generality of IQ effects across levels of cognitive development.

The developmental position described earlier predicts that when level of development is controlled through MA-matching, there should be no IQ-related differences in formal cognitive processes such as those involved in learning and problem solving. The difference position predicts that, with MA held constant, children of advanced IQ will be cognitively superior to children of average IQ, who will in turn be cognitively superior to retarded children.

In addition to testing earlier findings on IQ and hypothesis behavior, the present design permitted a reexamination of the finding (Weisz & Achenbach, 1975) that children of MA 5½ can use hypotheses. The Gholson et al. research cited above contradicts this finding, but because Gholson and McConville (1974) have demonstrated that kindergarteners given proper training can use hypotheses, and because children of MA 5½ in the study by Weisz and Achenbach verbalized numerous hypotheses, it was predicted that children at the 5½ year MA level, as well as those at higher levels, would use hypotheses in the present study.

One additional prediction was derived from evidence that retarded persons in relatively protective homogeneous school and community settings are likely to be more self-confident and more apt to use their own systematic problem-solving strategies than are retarded persons in less sheltered settings (cf. Achenbach & Zigler, 1968; Green & Zigler, 1962; Gruen, Ottinger, & Ollendick, 1974). Some retarded subjects in the present study were assigned primarily to homogeneous special-education classes, while others were assigned to regular classes for the nonretarded. The research of Zigler and his associates leads to the prediction that retarded children in regular classes will be less likely to use hypotheses than retarded children in special-education classes.

METHOD

Experimental Design

The 3 × 3 × 2 factorial design employed IQ levels of approximately 70, 100, and 130, crossed with MA levels of approximately 5½, 7½, and 9½ years. The third factor was mode of presentation of stimuli in the hypothesis-testing task. Half the subjects received stimulus cards in a stationary format on a table, while the remaining half received cards rotating on a turntable. Each of the 18 cells of the design contained five boys and five girls.

Subjects

To assess each child's IQ and MA, a male experimenter administered the Stanford–Binet short form, using an "optimizing" procedure by which difficult items were alternated with simpler ones (cf. Zigler & Butterfield, 1968).[1] Children at the different IQ and MA levels were assigned to the experimental conditions randomly with the constraint that no subgroup differences in mean IQ within an IQ level or in mean MA within an MA level be significant. MA and IQ were uncorrelated ($r < .08$).

The IQ range was 49 to 83 in the retarded group, 89 to 112 in the average group, and 118 to 145 in the bright group. The MA range was 55–79 months at the 5½-year level, 81–103 months at the 7½-year level, and 105–135 months at the 9½-year level. All children were white except for four black children, two in the low-MA retarded group, one in the middle-MA average

[1] The optimizing procedure, used in the earlier study by Weisz and Achenbach (1975), was designed to minimize adverse motivational influences on test performance and therefore to maximize the accuracy of IQ and MA scores as indexes of cognitive achievement. The use of more standard testing procedures, which appear to inhibit the expression of intelligence among children of lower SES (cf. Zigler & Butterfield, 1968), could have led to underestimates of the actual intelligence of such children. So, the optimizing procedure was used to provide for a more sensitive test of positive relationships between IQ and cognitive performance than would be obtained using more standard intelligence-test procedures.

group, and one in the high-MA retarded group. By examining school records, an effort was made to exclude all retarded subjects suffering from organic impairment. All subjects attended nursery school or public school in one of two small, predominantly middle-class cities. Those retarded children who were in special-education classes had been placed there because both IQ scores and school performance had indicated that such a placement was appropriate.

All seven steps on the Hollingshead SES scale were represented in the sample, but, as in most studies employing a broad range of IQ, SES and IQ were correlated ($r = -.44$, $p < .01$). In addition, largely because of the high social class of the preschoolers who made up the low-MA bright group, SES and MA were also correlated ($r = -.17$, $p < .05$). However, analyses of covariance with SES controlled revealed that group differences in SES affected neither the direction nor the level of significance of any analysis-of-variance finding reported in the present paper.

Procedure

Practice problems. Five to six weeks after administering the Binet, the experimenter again took each subject to the testing room and asked him/her to help "try out some new things for kids to do." The experimenter placed a deck of 20 white 7½ × 15-cm posterboard cards face up on a table between himself and the subject. One brown and one blue stimulus (square and star) appeared side by side on each card. Shapes were cut from colored sheets of Pantone, an acetate substance; on every card one shape was 6½ × 6½ cm, and the other was 2½ × 2½ cm. The letter *X*, 1 cm in height, was printed in the center of one shape; the letter *O*, also 1 cm in height, was in the center of the other shape. Thus, the two stimuli differed on four dimensions: shape, color, size, and letter. Within any specific blank-trials problem, the same two shapes, two colors, two sizes, and two letters were paired on each of the trials. In selecting the pair of shapes and the pair of colors to be used for a given problem, a pool of eight bisymmetrical shapes and eight colors was employed (with each shape and each color being used for 3 of the 12 problems); the same two letters and sizes were paired for every problem.

For Practice Problem 1 the subject was asked to name the two shapes, their colors, their relative sizes (bigger and smaller), and their letters. For subjects in the rotating condition, the deck of stimulus cards was then placed on a 15¼ × 7¼-cm white, Plexiglas platform with thin copper braces bounding the card deck at both ends. The platform was then placed on a turntable, 20 cm in diameter, rotating at 33⅓ rpm. For subjects in the stationary condition, the deck of cards remained in place on the table.

Practice Problems 1 and 2 were then administered, with feedback provided on every trial. For Practice Problem 3, feedback occurred on every second trial, and for Practice Problem 4, on every third trial. On each practice problem, the subject was informed on the first feedback-trial error occurring after six feedback trials:

> The right answer for this bunch of cards is one of the (the relevant dimension, e.g., *sizes*). Which, e.g., *size* do you think it is that's right every time? Okay, try picking (cue chosen by the subject, e.g.) *big*; see if *big* is the *size* that's right every time.

With each subsequent "wrong" feedback, the experimenter said, "Remember, the right answer is one of the *sizes*; which *size* do you think it is? . . . Okay, try picking the (*little*), etc." If the subject made the wrong choice after six additional feedback trials, she/he was informed of the correct answer. The experimenter then said, "Now let's see if you can be right every time," and proceeded through enough cards to provide a minimum of six additional feedback trials. On the first error after the sixth feedback-trial training stopped, the problem was recorded as having been failed, and the next problem was introduced. Only three subjects (two low-MA, superior-IQ; one low-MA, average) failed to attain the training criterion of six consecutive correct responses on three of the four practice problems; these subjects were replaced in order to complete the sample of 180. Unlike the experimental problems, practice problems did have correct answers rather than prearranged feedback; for the four practice problems the correct answers were a particular size, color, shape, and letter, respectively. At the end of each practice problem, the experimenter reviewed with the subject the correct answer and the fact that no other stimulus dimension was relevant.

Experimental problems. After Practice Problem 4 the subject was given two marbles and was told that she/he would be given a marble every time she/he was told "right," would forfeit a marble every time she/he was told "wrong," and would win a prize if she/he obtained seven marbles or more. After each experimental problem the subject checked his/her progress toward the goal of seven marbles, using a score sheet. Recent evidence (e.g., Harter & Zigler, 1972) indicates that a reward-plus-penalty procedure enhances motivation during discrimination learning.

At the beginning of every problem, all stimulus dimensions (e.g., color) were reviewed and all cues (e.g., brown, blue) were named. Each of the eight experimental problems consisted of 20 cards, with stimuli arranged as for the sample problems. In each problem it was possible to form eight stimulus pairs in which the paired stimuli differed along all four dimensions. In every problem, four such pairings were used only on feedback trials, and the remaining four appeared in four different randomly determined series, forming the blank-trial blocks. Each set of four stimulus cards was internally orthogonal in the sense that each level of every dimension appeared twice with each level of every other dimension. The arrangement provided that any consistent response to one cue of a stimulus dimension across a block of four blank trials (i.e., any legitimate hypothesis pattern) would produce a response pattern in the stationary condition which entailed three choices to one side of the card and one to the opposite side (cf. Gholson et al., 1972). This provision had the effect of disqualifying the three position-oriented response patterns which are most commonly reported among young children, i.e., position perseveration, alternation, and double alternation.

For the eight experimental problems, the experimenter said "right" or "wrong" only on Trials 5, 10, 15, and 20; thus, Trials 1 to 4, 6 to 9, 11 to 14, and 16 to 19 were blank-trial blocks. All feedback on the experimental problems was given according to the prearranged sequence used by Weisz and Achenbach (1975). Each time feedback occurred on practice and experimental problems, the experimenter pointed to the stimulus the subject had chosen and said, "You chose this one, and it was (right/wrong)," then silently counted off 4 sec before turning to the next card. On the final trial of six problems, children were told "right"; on the final trial of the remaining two problems, no feedback was given. This procedure ensured progress (with occasional regressions) toward the goal of seven marbles. After the final problem, each subject was informed that she/he had enough marbles to choose a prize, was asked to explain how she/he had done so well, and was then given a choice of several prizes.

RESULTS

Initially, each dependent measure was analyzed via a $3 \times 3 \times 2 \times 2$ (IQ $\times$ MA $\times$ mode of presentation $\times$ sex) ANOVA; however, only 1 of the 32 F tests involving effects of sex was significant. Thus, in the service of clarity, analyses reported below are in the form of $3 \times 3 \times 2$ (MA $\times$ IQ $\times$ mode of presentation) ANOVAs.

Graded Help

During each of the four practice blank-trials problems, as many as two steps of graded help, or hints, were given if needed (see description in Method section). The number of hints given across all four problems served as a rough index of the difficulty encountered by each child in comprehending the procedure. Fewer hints were required with increasing MA [$F(2,160) = 16.15, p < .001$]; the linear component of this MA effect was highly significant [$F(1,160) = 32.41, p < .001$]. In addition, fewer hints were required for problems in the stationary condition than for those in which the stimuli rotated [$F(1,160) = 5.17, p < .05$].

Reliance on Hypotheses

In scoring hypothesis behavior, those response patterns across a blank-trial block which were consistent with one of the simple, nonposition solutions outlined for subjects were designated hypotheses. Subjects responding randomly would manifest hypothesis patterns on an average of 16 of the 32 blank-trial blocks; thus, the prediction that all groups would use hypotheses was tested by comparing each of the 18 subgroup means shown in Table 1 to the chance baseline of 16. Each of these means was above this chance baseline, and the mean for every group except one (i.e., retarded subjects at the middle MA level who received stationary stimuli) was significantly greater than 16; in each instance $t\ (9) > 2.1, p < .05$ (all tests one-tailed).

The only statistically significant effect on number-of-hypothesis patterns was that of MA level (means: 22.3, 22.6, 25.4) [$F(2,162) = 5.33$, $p < .01$], and the linear component of this effect was significant [$F(1,162) = 8.07$, $p < .01$]. Although the MA × mode of presentation interaction did not approach significance, it should be noted that at the two lowest MA levels, where position-oriented response sets are often reported, children using the rotating stimuli designed to discourage position responding did manifest more hypothesis patterns than children in the stationary condition where position-oriented responding was merely disallowed (means: 22.9 vs 21.6 at the low MA level, 23.8 vs 21.4 at the middle MA level).

Position Responding

Among children in the stationary-stimulus condition, it was possible to note three common types of position responding. (This was, of course,

TABLE 1

NUMBER OF HYPOTHESIS PATTERNS: MEANS AND CONFIDENCE INTERVALS[a] FOR EACH EXPERIMENTAL GROUP

	Rotating stimuli			Stationary stimuli		
MA Level	Retarded	Average	Bright	Retarded	Average	Bright
5½ Years						
Mean	22.70	23.70	22.40	22.10	20.80	21.90
Confidence interval	±3.86	±3.41	±3.00	±3.53	±3.57	±5.76
7½ Years						
Mean	23.80	23.80	23.80	20.30	20.60	23.40
Confidence interval	±2.20	±4.59	±4.64	±6.78	±4.95	±4.22
9½ Years						
Mean	21.40	25.40	28.60	23.80	25.00	28.30
Confidence interval	±3.51	±3.47	±2.70	±3.85	±4.26	±3.74

Note. Expected mean with random responding (i.e., chance baseline) = 16.

[a] Based on two-tailed t value at $p = .05$.

impossible in the rotating condition where stimulus position was constantly changing.) Mean cell-by-cell frequencies of position perseveration (e.g., RRRR), position alternation (e.g., RLRL), and double alternation (e.g., RRLL) are shown in Table 2. As might be expected from the ANOVA results with number-of-hypothesis patterns, the total number of position-oriented nonhypothesis patterns was significantly related to MA [$F(2,81) = 3.91, p < .05$]. As MA level increased, the number of position-oriented response patterns on the blank-trial blocks decreased (means: 9.0, 8.1, 4.9). To determine which specific type(s) of position responding might account for this MA effect on total position responses, the three component measures shown in Table 2 were analyzed by means of 3 × 3 (MA × IQ) ANOVAs. The only significant effect was the main effect of MA on number-of-position perseveration patterns [$F(2,81) = 5.63, p < .01$]. As was the case with total position responses, the number of position perseveration patterns declined with increasing MA (means: 3.8, 1.8, 1.2).

Responsiveness to "Right"

The next analysis concerned the manner in which subjects used hypotheses on those blank-trials blocks which followed the feedback information "right." The measure used was the arcsine transformed percentage of times a hypothesis pattern followed by the feedback "right" led to repetition of the same hypothesis pattern on the next set of four blank trials. On repetitions following "right," there was a significant IQ × mode of presentation interaction [$F(2,160) = 6.10, p < .01$]; this reflected the fact that retarded subjects repeated more often with stationary than with

TABLE 2

NUMBER OF POSITION-ORIENTED RESPONSE PATTERNS: MEANS FOR GROUPS RECEIVING STATIONARY STIMULI

MA Level	Retarded	Average	Bright
5½ Years			
Position perseveration	2.3	4.3	4.8
Position alternation	3.5	4.6	2.9
Double alternation	2.6	0.9	1.0
7½ Years			
Position perseveration	1.1	1.4	2.8
Position alternation	6.7	5.9	2.3
Double alternation	0.9	1.5	1.6
9½ Years			
Position perseveration	1.2	1.1	1.2
Position alternation	2.8	2.8	1.1
Double alternation	2.0	1.7	0.7

rotating stimuli, while average and superior subjects repeated more often when stimuli were rotating. In addition, there was an interaction of MA and IQ [$F(4,160) = 3.08$, $p < .05$]; the interaction reflects distinct changes with MA level in the relative status of the three IQ groups (see Table 1). Only at the 9½-year MA level did the retarded score lowest and superior subjects highest. Separate 3 × 2 (IQ × mode of presentation) ANOVAs at each MA level revealed that the main effect of IQ was nonsignificant at the two lower MA levels, but significant as the highest MA level [$F(2.54) = 3.72$, $p < .05$].

Responsiveness to "Wrong"

The next analysis was concerned with the way subjects used hypotheses on those blank-trial blocks which followed the feedback information "wrong". The measure used for this purpose was the arcsine transformed percentage of times a hypothesis pattern followed by the feedback "wrong" led to a change to any other hypothesis pattern. As shown in Table 3, changes in hypotheses following "wrong" increased significantly with MA [$F(2,161) = 5.32$, $p < .01$]; the linear trend of MA was significant [$F(1,161) = 9.85$, $p < .01$].

Effects of Classroom Environment

A final purpose of this study was to test the prediction that retarded persons who attend class with nonretarded peers are less likely to use strategies or hypotheses during learning than are retarded persons whose

TABLE 3

MEAN PERCENTAGE OF HYPOTHESIS PATTERNS REPEATED FOLLOWING "RIGHT" AND REPLACED FOLLOWING "WRONG"

MA Level	Retarded	Average	Bright
5½ Years			
Right	63	66	53
Wrong	44	55	35
7½ Years			
Right	64	41	66
Wrong	53	54	56
9½ Years			
Right	56	70	82
Wrong	45	69	77

Note. The figures after Wrong represent the mean percentages of times that a hypothesis pattern and the subsequent feedback "wrong" were followed by a different hypothesis pattern.

milieu is designed specifically for the retarded. This test was difficult to structure because most retarded subjects at the two lower MA levels sampled were pupils in full-time special-education classes, while nearly all upper-MA retarded subjects were required, by school policy, to spend at least half of every school day in classes with nonretarded pupils. It was possible, however, to form 11 pairs of retarded subjects, matched on MA and mode of stimulus presentation, but differing in that one member was identified primarily with a special-education class while the other member was identified primarily with a regular class for the nonretarded. The special-class and regular-class groups did not differ significantly in gender or racial composition or in SES. There was a distinct bias against the hypothesis being tested, in that the special-education subjects were consistently lower in IQ than their regular class cohorts (means: 64.3 and 77.4) [$t(10) = 6.93, p < .001$], and the children had been placed in special classes originally because of both low IQ scores and poor academic performance. Despite this bias, the special-education subjects, compared to those from regular classes, had significantly higher mean scores on number-of-hypothesis patterns [$t(10) = 2.36$, $p < .05$ (one-tailed test)]. In addition, the special-education group scored nonsignificantly higher than the regular-class group on both measures of responsiveness to feedback.

DISCUSSION

Taken as a whole, the results of this investigation support the *developmental* position that, regardless of developmental rate, similarity of developmental level implies similarity of cognitive competence (cf. Zigler, 1969). On no hypothesis measure was there a significant main effect of IQ. Responsiveness to the feedback "right" was positively related to IQ only at the 9½-year MA level, not at the two lower MA levels; this inconsistency of IQ effects across MA levels is not consonant with the *difference* view that individuals of similar MA but different IQ are inherently different in intellectual ability. Weisz and Achenbach (1975) found that on four-dimensional problems in which one dimension was position, retarded subjects used fewer hypothesis patterns and were less responsive to the information "right" than nonretarded subjects of similar MA; the present data indicate that such "IQ effects" may be a function of the experimental methodology employed rather than a reflection of some cognitive deficit invariably associated with low IQ. The findings clearly demonstrate the need for a sequential approach to the developmental vs difference controversy, an approach by which apparent effects of IQ are tested for their stability across significant variations in method.

The need for a *series* of studies of this genre is further underscored by the fact that Zigler's developmental prediction of no IQ-group differences is favored by conventional significance testing procedures. For this reason substantial *cumulative* support would be required for confirmation of the developmental theory. Thus, while the present findings may be regarded as harmonious with and supportive of the theory, alone they are in no respect conclusive. On the other hand, it should be noted that the experimental control in the present study was sufficient to permit the detection of MA effects, and this fact supports the conclusion that the absence of an IQ effect was not simply an artifact of excessive variability in the data.

A procedure using rotation of stimuli to discourage position responding proved slightly more difficult for subjects to grasp in pretraining than a procedure in which stimuli were stationary, but subjects in the two lower MA levels who received rotating stimuli used nonsignificantly more hypothesis patterns than subjects of similar MA whose stimuli were stationary. This suggests that the rotating procedure may have had some of its intended effect in discouraging stereotyped, nonhypothesis responding. Subjects using rotating stimuli repeated hypotheses following the feedback information "right" somewhat less often than did subjects in the stationary condition, and this may indicate that the rotation of stimuli made it more difficult to track specific stimulus cues across blocks of blank trials. It is possible that such a disadvantage in tracking specific cues offset any advantage the rotating procedure may have had as a means of discouraging position-oriented responding. But, for whatever reason, mode of stimulus presentation had no significant main effect on any hypothesis measure other than graded help.

As was found by Weisz and Achenbach (1975), children at the 5½-year MA level, as well as those of greater maturity, used hypotheses. However, the percentages of blank-trial blocks on which low-MA groups manifested hypothesis patterns in the present study (range, 65–74%) were lower than the corresponding figures reported by Weisz and Achenbach (range, 75–84%) on four-dimensional problems. In addition, repetitions of hypothesis patterns following the feedback information "right" were more numerous in the previous study than in the present one; and subjects in the previous study were less effective at switching hypothesis patterns following "wrong" than in the present investigation. Taken together, these facts suggest that low-MA subjects in the Weisz–Achenbach sample may have engaged in some repetitive responding, which interfered with their responsiveness to the feedback information "wrong," and that such responding is discouraged when position is eliminated from the pool of legitimate hypothesis dimensions. The evidence does suggest that subjects of kindergarten MA levels can generate hypotheses on blank-trial problems if they are given thorough pretraining in the differentiation and use of stimulus dimensions.

In the present sample, the tendency to retain a hypothesis after the feedback "right" varied directly with IQ only as MA increased. This interaction of MA and IQ is reminiscent of earlier findings that differences in academic achievement between middle- and lower-class children become greater as the children spend more years in school (Bartel, 1971; Jensen, 1966). Research into learned helplessness (cf. Seligman, 1974) and expectancy of failure (cf. Zigler, 1971) indicates that debilitating effects on problem solving of a perceived inability to control outcomes can be traced to repeated experiences in which individuals fail to effect the outcomes they desire. The cumulative total of such failure experiences may not differ greatly among bright, average, and retarded children at low levels of maturity (e.g., MA 5½ years) where little formal schooling has yet occurred; but at levels of maturity (e.g., MA 9½ years), where children have had some exposure to the competitive rigors of classroom experience, it is likely that the experience of failure will be most common among retarded children and least common among bright children. This seems especially likely where special-education classes are not available. The small sample comparisons made in this study suggest that deficiencies in hypothesis behavior may be more pronounced among retarded children in regular classrooms than among those placed in special-education classes. This finding, consistent with earlier reports on the beneficial effects of protective settings in which behavioral expectations are tailored to the abilities of the child (e.g., Green & Zigler, 1962; Gruen et al., 1974), raises an important issue for further study within the current controversy over "mainstreaming" versus special-class placement of retarded children (cf. MacMillan, Jones, & Aloia, 1974).

REFERENCES

Achenbach, T. M., & Zigler, E. Cue-learning and problem-learning strategies in normal and retarded children. *Child Development*, 1968, **39**, 827–848.

Bartel, N. R. Locus of control and achievement in middle- and lower-class children. *Child Development*, 1971, **42**, 1099–1107.

Denny, M. R. Research in learning and performance. In H. A. Stevens & R. Heber (Eds.), *Mental retardation: A review of research*. Chicago: Univ. of Chicago Press, 1964. Pp. 100–136.

Eimas, P. D. A developmental study of hypothesis behavior and focusing. *Journal of Experimental Child Psychology*, 1969, **8**, 160–172.

Gholson, B., Levine, M., & Phillips, S. Hypotheses, strategies, and stereotypes in discrimination learning. *Journal of Experimental Child Psychology*, 1972, **13**, 423–446.

Gholson, B., & McConville, K. Effects of stimulus differentiation training upon hypotheses, strategies, and stereotypes in discrimination learning among kindergarten children. *Journal of Experimental Child Psychology*, 1974, **18**, 81–97.

Green, C., & Zigler, E. Social deprivation and the performance of retarded and normal children on a satiation type task. *Child Development*, 1962, **33**, 499–508.

Gruen, G. E., Ottinger, D. R., & Ollendick, T. H. Probability learning in retarded children with differing histories of success and failure in school. *American Journal of Mental Deficiency*, 1974, **79**, 417–423.

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., & Zigler, E. Effects of rate of stimulus presentation and penalty conditions on the discrimination learning of normal and retarded children. *Developmental Psychology*, 1972, **6**, 85–91.

Heal, L. W., & Johnson, J. T. Inhibition deficits in retardate learning. In N. R. Ellis (Ed.), *International review of research in mental retardation, Vol. IV*. New York: Academic Press, 1970. Pp. 107–149.

Jensen, A. R. Cumulative deficit in compensatory education. *Journal of School Psychology*, 1966, **4**, 37–47.

Levine, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–338.

Macmillan, D. L., Jones, R. L., & Aloia, G. F. The mentally retarded label: A theoretical analysis and review of research. *American Journal of Mental Deficiency*, 1974, **79**, 241–261.

Seligman, M. E. P. *Helplessness*. San Francisco: W. H. Freeman, 1974.

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

Zeaman, D. One programmatic approach to retardation. In D. K. Routh (Ed.), *The Experimental Psychology of Mental Retardation*. Chicago: Aldine, 1973. Pp. 78–132.

Zeaman, D., & House, B. J. The relation of IQ and learning. In R. Gagné (Ed.), *Learning and individual differences*. Columbus, Ohio: Charles E. Merrill Books, 1967.

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

REFERENCE NOTES

1. Hollingshead, A. B. *Two-factor index of social position*. Unpublished manuscript, Yale University, 1957.
2. Weisz, J. R., & Zigler, E. Current status of the developmental versus difference controversy in the area of mental retardation. In preparation.

RECEIVED: July 9, 1976; REVISED: September 27, 1976