

Selective Screening Device for the Early Detection of Normal or Delayed Cognitive Development in Infants at Risk for Later Mental Retardation

Joseph F. Fagan III, PhD, Lynn T. Singer, PhD, Jeanne E. Montie, MA, and Patricia A. Shepherd, PhD

From the Department of Psychology, Case Western Reserve University, Cleveland

ABSTRACT. The present study tested the predictive validity at 3 years of age of a screening device for the early identification of later cognitive delay. The screening device, administered between 3 and 7 months of age, is based on the infant's differential fixation "to novel" over previously shown pictures. The sample was composed of 62 infants suspected to be at risk for later mental retardation. The prevalence of delayed cognitive development (IQ \leq 70) at 3 years of age was 13%. Novelty preference scores correctly identified six of eight (75%) of the delayed children. The test identified 49 of 54 (91%) of the normal children. Validity for predicting cognitive delay was 55%. Validity for the prediction of normality was 96%. The screening device proved to be equally sensitive, specific, and valid when the sample was divided into infants born at term or born preterm. The results of the present study and of a previous study indicate that detection of cognitive delay based on early novelty preferences is as easily accomplished for infants who will later be mildly delayed (IQ scores 60 to 70) as it is for those who will later be severely delayed (IQ scores \leq 50). Moreover, such results are in contrast to those obtained with conventional tests tapping sensorimotor development. *Pediatrics* 1986;78:1021-1026; *retardation, screening, cognition, memory, novelty.*

The present study presents initial data on the predictive validity at 3 years of age of a screening device for the early identification of later cognitive delay. The screening device, which is administered between 3 and 7 months of age, is based on the infant's ability to recognize visual stimuli. Visual recognition is measured by differential fixation "to novel" over previously shown pictures.¹ Differential

fixation to novel over previously shown pictures demonstrates that the infant can discriminate among the pictures being shown and can remember the picture previously seen. In other words, tests of visual novelty preference yield information on the infant's developing ability to perceive and to retain information. To the extent that perceptual and memory processes are necessary for successful performance on later intelligence tests, it seems reasonable to suppose that individual differences in recognition memory during infancy would be related to individual differences in later intelligence. In fact, individual differences in novelty preferences during infancy have been shown to be predictive of later intelligent functioning.²

Given that early novelty preferences are linked to later intelligence, a major goal of our research has been to develop a screening device for later intellectual deficit based on early novelty preferences.³ The purpose of the screening device is to differentiate potentially normal from potentially deficient infants within groups of infants suspected to be at risk for later mental retardation. Suspicion of risk would be due to the presence of various prenatal or perinatal factors (eg, extreme prematurity).

The rationale for choosing novelty preferences as the metric for infant intelligence and details on the construction of, normative data on, and the reliability of the screening device have been discussed previously.³ Briefly, the basic component of the screening procedure is a "novelty problem." Each novelty problem consists of a pairing of two stimuli immediately following a standard period of study to one of the two stimuli.

Our approach to the development of a practical standard assessment of individual differences in

Received for publication June 27, 1985; accepted Feb 26, 1986.
Reprint requests to (J.F.F.) Department of Psychology, Case Western Reserve University, Cleveland, OH 44106
PEDIATRICS (ISSN 0031 4005). Copyright © 1986 by the American Academy of Pediatrics.

early recognition memory was to allow infants a particular length of study time, a length somewhat less than necessary for asymptotic novelty preferences for the average infant of that age, and then to test for recognition by pairing a novel and previously seen target.³ Following this procedure with a variety of novelty problems allowed each infant a composite novelty preference score derived from many items. Standard assessment meant that all infants at a particular age receive the same targets as novel or familiar, which necessitated initial studies with 180 infants in which preferences within pairs of targets were obtained. The second step in this development was to discover which combinations of standard study times (ie, all infants were allowed the same study time on a particular problem) and target pairings would lead, at various ages, to a set of novelty problems that could be easily administered during a single session and which, individually, could be solved (ie, preferences for novelty greater than chance for each problem). The second step was undertaken in a series of cross-sectional and longitudinal studies involving 1129 testing sessions. The result of all this effort was a screening device composed of 12 novelty problems, three of which are administered at 52, two at 56, four at 62, and three at 69 weeks of conceptional age (ie, for term infants at 12, 16, 22, and 29 weeks of postnatal age). A description of the pictures used at each age and the details of test administration have been published previously.³ The next step in test development was to develop norms for the 12-item test by conducting a longitudinal study using a sample of 92 infants with recognition tests administered at 52, 56, 62, and 69 weeks of conceptional age. Infants in the normative sample showed significant visual preferences for novelty on each of the 12 novelty problems. Finally, the screening device, known as the Fagan Test of Infant Intelligence, proved to have concurrent validity because groups of infants at risk for later intellectual deficit in the study by Fagan and Singer³ performed less well on the test at 3 to 7 months of age than did the 92 normal controls.

The present study followed the same high-risk infants originally tested by Fagan and Singer³ to 3 years of age in order to estimate the predictive validity of the screening device. Specifically, we wanted to know if children who did poorly on the Fagan test as infants would also exhibit delayed cognitive development at 3 years of age. Delayed cognitive development at 3 years of age was operationally defined as an IQ score of 70 or less on a standard test of intelligence. Although the predictions from the infant tests have only been followed to 3 years of age, thus far, it is highly likely that

the distinctions between normal and delayed cognitive development made at 3 years will still be valid at school age. Wilson,⁴ for example, has shown with sample sizes of more than 400 children that even within normal populations IQ scores remain relatively stable from initial intelligence tests at 3 years of age to later intelligence tests at ages 6 to 9 years (test-retest coefficients of .65 to .73).

SUBJECTS AND METHODS

The sample was composed of 62 infants suspected to be at risk for later mental retardation as a result of prematurity (birth weight <1,500 g), intrauterine growth retardation, treated hypothyroidism, a diagnosis of failure to thrive, or a history of maternal diabetes. The conditions predisposing to risk will not be considered further here. Empirically, we simply hoped to draw infants from populations³ in which the incidence of IQ scores ≤ 70 at 3 years of age was expected to be greater than 2%. All infants were tested at home by means of a portable apparatus.¹ The apparatus included a vertical screen upon which two 18 × 18-cm targets were placed 30 cm apart from the center of one target to the center of the other and 30 cm from the infant's eye. Through a 0.64-cm peephole in the center of the screen, an observer could see corneal reflections of the targets. Length of superimposition of the target over the pupils was recorded by means of a finger switch that activated a digital recorder. Interobserver reliability in the measurement of infant differential fixation is typically >.90.⁵

As noted, the screening device administered to each infant was composed of 12 novelty problems, three of which were administered at 52, two at 56, four at 62, and three at 69 weeks of conceptional age. Specifically, the 12-problem novelty test included three pairs of abstract, black and white patterns that differed from one another along a variety of dimensions constituting three novelty problems at 12 weeks. Study times of 60 seconds were allowed for each problem given at 12 weeks, ie, the infant looked for 60 seconds at a target before it was paired with a new picture. At 16 weeks, two pairs of abstract black and white patterns were used as novelty problems. One pair varied along a variety of dimensions and infants were only allowed 16 seconds to study one of the pictures in that pair prior to its pairing with the other (novel) member of the pair. The second problem at 16 weeks contained two targets differing solely in arrangement of internal pattern elements. Infants were allowed 60 seconds of study for the second problem at 16 weeks. At 22 weeks, two sets of achromatic face photos (man *v* woman, woman *v* baby) with study times of 40 seconds and two pairs of abstract pat-

terns varying in arrangement of elements with study times of 30 seconds were used. Study times decreased to 20 seconds per problem at 29 weeks when three pairs of face photos were used, including two pairs with achromatic photos of women and one pair of chromatic prints of babies' faces.

Every attempt was made to test each infant at each of the four ages (12, 16, 22, and 29 weeks of corrected age), but more infants were seen at 22 or 29 weeks than were seen at 12 or 16 weeks due to late referrals. In the final analysis, infants were only included in the sample if they had received at least seven of the novelty problems. Such a restriction ensured that an infant was seen at least twice with at least one test given at 22 or 29 weeks.

The score assigned to each infant on the Fagan test was a mean novelty preference score. For example, if an infant were shown a new and a previously seen pattern and devoted 10 seconds of looking time to the new pattern and six seconds to the old pattern, his or her percentage of total fixation to the new pattern would be 10/16 or 63% for that problem. The sum of these percentages to novelty divided by the number of problems received yielded a mean novelty preference score for each infant.

Intellectual outcome at 3 years of age was assessed by administering the Stanford-Binet⁶ and the Peabody Picture Vocabulary Test (Form L)⁷ and giving the child a mean IQ score for the two tests. If necessary, the Bayley Scales of Mental Development⁸ were administered and an IQ score was estimated by dividing the age attained on the test by chronologic age. The technicians who administered the intelligence tests to the 3-year-old children were unaware of how the child had performed as an infant.

RESULTS AND DISCUSSION

The sample yielded a mean novelty preference score of 59.5% (SD 8.1) and a mean IQ at 3 years of age of 96.3 (SD 23.1, range 25 to 135). Delayed cognitive development at 3 years was defined as an IQ score of 70 or less. The incidence of cognitive

delay (IQ \leq 70) at 3 years was 13%. Novelty preference scores of \leq 53% correctly identified six of eight (75%) of the delayed children. The children correctly identified had IQ scores of 25, 25, 60, 70, 70, and 70. The two children predicted to be normal who were ultimately found to be delayed had IQ scores of 25 and 68. The child with the IQ of 25 had a median novelty preference score of $<$ 53% which would have identified her as at risk. Her mean novelty preference score was 56% due to a high percentage on one problem on an early test. She had performed particularly poorly on the 5- and 7-month novelty problems. The other child had a mean novelty preference score of 70%. There was no evidence to indicate that any of the eight cognitively delayed children had visual deficits that would have impaired their performance on the novelty problems. All perinatal conditions associated with the eight delayed children were also represented among the 54 children whose IQ scores were $>$ 70 at 3 years of age.

High sensitivity in predicting delayed cognitive development was not accomplished at great cost to specificity for normality, because the test also correctly identified 49 of 54 or 91% of the normal children. In the present sample, the validity of the screening device for predicting cognitive delay was 55%, because six of 11 infants predicted at 7 months of age to be at risk were indeed delayed at 3 years of age. Validity for the prediction of normality was 96%, ie, 49 of 51 predicted to be normal were later found to be normal. As one can see from the data presented in Table 1, the screening device proved to be equally sensitive, specific, and valid when the sample was divided into those infants born at term (gestation $>$ 36 weeks) and those born preterm.

In an earlier study,⁹ a preliminary version of the Fagan Test of Infant Intelligence consisting of four novelty problems was administered between 5 and 7 months of age to a group of 20 infants diagnosed as failure to thrive. The incidence of delayed cognitive development at 3 years of age was 40% (eight of 20 had IQ scores $<$ 70). Novelty preference scores

TABLE 1. Sensitivity and Specificity of Tests for Visual Novelty Preference Given at 3 to 7 Months of Age for the Prediction of Cognitive Delay at 3 Years of Age

	Intellectual Outcome at 3 Yr of Age					
	All Children (n = 62)		Term Children (n = 34)		Preterm Children (n = 28)	
	Delayed	Normal	Delayed	Normal	Delayed	Normal
Risk prediction	6	5	3	2	3	3
Normal prediction	2	49	1	28	1	21
Sensitivity to delay	.75		.75		.75	
Specificity to normality	.91		.93		.88	
Predictive validity for delay	.55		.60		.50	
Predictive validity for normality	.96		.97		.95	

correctly identified seven of eight of the children with IQ scores ≤ 70 for a sensitivity of 88% and ten of 12 of the normal children for a specificity of 83%. The validity for predicting delayed development was 78% (seven of nine predicted to be at risk for delay were later found to be delayed). The validity for the prediction of normality was 91% (ten of 11 correct predictions). Estimates of sensitivity, specificity, and prediction for normality are similar from the earlier to the present study. The disparity, from the earlier to the present study, between values for the prediction of cognitive delay (78% and 55%, respectively) must be qualified by the fact that such estimates will vary with the prevalence of positive cases. It is to be expected that a higher validity for the prediction of delayed development would be associated with a higher prevalence of cognitive delay.

If the data yielded by the 62 children in the present study and the 20 failure to thrive children studied by Fagan et al⁹ are combined, it is clear that the visual novelty test is as predictive for children with mild and moderate intellectual delay (IQ scores 50 to 70) as it is for children afflicted with severe or profound intellectual delay (IQ scores < 50). Of the 16 children (from both studies) found to be delayed at 3 years of age, four had IQ scores < 50 , one had an IQ score of 50, and 11 had IQ scores between 60 and 70. The visual novelty test correctly predicted cognitive delay for three of four of the children with IQ scores < 50 and correctly identified as delayed ten of the 12 children with IQ scores from 50 to 70.

In addition, combining the present with the previous study⁹ makes it possible to compare directly the relative value of visual novelty tests and conventional sensorimotor tests of infant "intelligence" for the prediction of later cognitive delay. Twenty-seven of the 82 children in the two samples combined had also been assessed with the Bayley scales during infancy. All 27 children had, as infants, been diagnosed as failure to thrive. The Bayley scales were given to each infant by clinical psychologists at various health care facilities as part

of a standard testing procedure. The 27 children given both types of assessment had a mean IQ, at 3 years of age, of 74.4 (SD 20.2, range 25 to 115). Prediction from early novelty preference scores for this subsample of 27 children was highly sensitive, with 91% (or ten of 11) of the children correctly identified as delayed, and was specific in identifying 81% (13 of 16) as normal at 3 years of age (Table 2). Predictive validity estimates for intellectual delay and for normality on the novelty preference test were high at 77% and 93%, respectively. In contrast, the Bayley scales administered between 6 and 8 months of age were low in sensitivity, with Mental Development Index scores ≤ 80 correctly identifying only 45% (five of 11) of the delayed children. The scales were also low in specificity because they correctly identified only 38% (six of 16) of the normal children. The validity of the Bayley scores administered when the infants were 6 to 8 months of age for the prediction of cognitive delay or normality was correspondingly low at 33% and 50%.

Our finding that the Bayley scales were low in predictive validity is not unusual. Infant "mental" tests based on sensorimotor functioning such as the Bayley scales have proven to be ineffective in predicting later intelligence among normal populations.³ The predictive power of such tests is often assumed to be much higher for infants who score low on the examinations. However, a review by Kopp and McCall¹⁰ demonstrated that, except for some specific syndromes, predictions for high risk and clinic samples can still not be considered practically useful. Fagan and Singer,³ for example, in reviewing studies of the predictive validity of infant sensorimotor tests such as the Bayley scales noted that correlations obtained between tests given during the 3- to 7-month period and standard IQ tests at 3 years of age or beyond average about .18 for high-risk and clinic samples. This is not to say that sensorimotor tests are not useful. Such tests do help investigators identify deficits in sensorimotor functioning. But as predictors of later cognitive performance, such infant tests have lacked accuracy.

TABLE 2. Sensitivity, Specificity, and Validity of Tests for Visual Novelty Preference Given at 3 to 7 Months of Age and Bayley Mental Development Index Given at 8 Months of Age for the Prediction of Cognitive Delay at 3 Years of Age (N = 27)

	Visual Novelty		Bayley Mental Development Index	
	Delayed	Normal	Delayed	Normal
Risk prediction	10	3	5	10
Normal prediction	1	13	6	6
Sensitivity to delay		.91		.45
Specificity to normality		.81		.38
Predictive validity for delay		.77		.33
Predictive validity for normality		.93		.50

In brief, tests of visual novelty preference administered to infants between 3 and 7 months of age appear to constitute a valid screening device for the early detection of intellectual deficit in high-risk populations. Moreover, detection appears to be as easily accomplished for infants born preterm as for those born at term and for infants who will later be mildly delayed (IQ scores 60 to 70) as it is for those who will later be severely delayed (IQ scores 50 and below). Such results are in contrast to those obtained with conventional tests tapping sensorimotor development such as the Bayley Scales of Mental Development.

IMPLICATIONS

As the results of this study indicate, a screening device based on early novelty preferences predicts with some accuracy which infants within samples of high-risk infants will later exhibit delayed cognitive development and which will later be normal. The development of such a technology raises the question of whether any benefits would accrue by transferring the technology to the clinician or to the clinical researcher. The benefit of screening for a disorder is usually defined in terms of an effective treatment program which prevents the damage that would have resulted had the condition gone undetected. Unfortunately, existing treatment programs are limited in their effectiveness for those children who would be detected by the visual novelty preference test to be at risk for mental retardation.¹¹ Thus, the benefits to use of the screening device currently lie in areas other than the sure remediation of positive cases.

The development of a valid screening device for differentiating infants who will later be intellectually normal from those who will later be delayed within a group of high-risk infants has four main benefits. The first benefit flows from two facts. One is that the validity for the prediction of normality based on the screening device is quite high. The second fact is that the incidence of later retardation among high-risk infants, although higher than in a normal population, is, on an absolute basis, usually low (eg, 13% of the present sample). The first benefit, then, is that parents of the majority of high-risk infants screened (about 90%) could be encouraged to view their infants as highly likely to be intellectually normal and would be spared the uncertainty and anxiety of waiting 2 or 3 years for a valid intellectual assessment. Such relief is a considerable sum in emotional currency.

The second benefit is that the cost of further testing and intervention can be greatly reduced by allocating scarce resources only to the infants who score positive on the test. It is important to empha-

size that infants suspected to show impaired cognitive development should not be labeled as retarded. Rather, they would be referred for more extensive follow-up testing. If the children continue to show poor performance with time, their families could be given needed emotional support and guided toward early intervention programs. As with any screening device, one potential difficulty with the Fagan test is that some children tested as infants will have false positive results. That is, children who will ultimately have normal intelligence will be considered at risk for retardation on the basis of their scores on the infant test (eg, 9% in the present study). The problem of labeling normal children as at-risk is attenuated by two facts. First, the Fagan test has been developed for use with populations already considered to be at risk for intellectual deficit, eg, infants with very low birth weights, intrauterine growth retardation, seizures. Presumably, parents of all such infants are anxious about the development of their infant. Infants within these high-risk populations receiving a test score that places them at risk for retardation will simply be followed more closely. In other words, parental anxiety need not be increased. Second, the Fagan test is not meant to be used as the sole criterion for a designation of intellectual risk. The test should be used in conjunction with other global infant assessments and the clinical judgment of specialized personnel (eg, neurologists, physical and speech therapists) where appropriate, in addition to regular pediatric care. Decisions regarding an infant's level of risk for intellectual deficit should be made after careful consideration of all possible input.

The third benefit of a valid screening device is that the early identification of mental retardation might allow a fuller understanding of the causes of intellectual deficit. Currently, we are unaware of the factors underlying the majority of cases of moderate to mild mental retardation. One reason for this lack of knowledge is that years must pass before mental retardation in the 50- to 70-IQ range can be diagnosed. Hence, we do not know which of a myriad of factors over the years singly or in combination might have led to cognitive malfunction. Use of visual novelty tests to detect infants at risk for retardation means that the lead time for longitudinal prospective studies of conditions predisposing to risk for mental retardation can be cut from 3 or more years to 1 year or less. Thus, the study of early novelty preferences may be used to pinpoint quickly those factors associated with conditions such as low birth weight, fetal and infant growth retardation, prenatal exposure to potential teratogens, infections, intraventricular hemorrhage, hy-

poglycemia, diabetic pregnancy, neonatal seizures, etc, which are most likely to be associated with early cognitive impairment. In the future, researchers using the screen for early detection of retardation may discover the etiology of certain cases of mental retardation (eg, by identifying behavioral teratogens) and may institute programs to prevent retardation in specific cases. Even a very limited success in research in the etiology and prevention of mental retardation would be of social benefit.

The fourth major benefit of a valid screening device for the early detection of retardation is its use in providing an assessment of the effects of intervention. Obstetricians might wish to determine the cognitive sequelae of in vitro fertilization, of drugs used to initiate labor or used to prevent premature birth, or of fetal diagnostic procedures. Pediatricians are concerned with evaluating the effects of feeding formulas and vitamin supplements and with the efficacy of various mechanical treatment procedures. Presently, changes in fetal and perinatal care occur rapidly. As noted, intellectual status can only be evaluated years after initial treatment. A screening device for early deficits in intelligence would fill the gap in time between treatment and assessment.

In brief, we have the beginnings of a valid selective screening device for the early detection of later mental retardation. The selective screening device is based on the infant's preference for visual novelty. As noted in the introduction, many years have been spent and more than 2,000 visits have been made to infants to decide, empirically, which novelty problems to include in the Fagan test. It is important to note that not just any pairing of a briefly studied picture with a novel one will differentiate normal from delayed infants or predict later intelligence. The choice of pictures and the time given for study of the pictures determines whether or not a novelty preference will be useful in differentiation and prediction. At this point, we believe that it may be beneficial socially to begin the process of transferring the screening device to the medical community. In suggesting such technology transfer, we cannot emphasize too strongly that the Fagan test, in its current form, should not be used for routine screening with normal populations. The

Fagan test should only be used as a selective screening device, ie, one that is applied only after an infant is suspected to be at risk for later cognitive deficit. The Fagan test is not intended for use by clinicians in primary care practice. Rather, the test, as currently developed, may be most beneficial to the clinician working with high-risk infants in a secondary or tertiary care center. In particular, the Fagan test should be most useful to the clinical researcher.

ACKNOWLEDGMENT

This research was supported, in part, by a Perinatal Emphasis Research Center grant (HD-11089) and by a Mental Retardation Research Training grant (HD-07176), both from the National Institute of Child Health and Human Development.

REFERENCES

1. Fagan JF: Memory in the infant. *J Exp Child Psychol* 1970;9:217-226
2. Fagan JF: The intelligent infant: theoretical implications. *Intelligence* 1984;8:1-9
3. Fagan JF, Singer LT: Infant recognition memory as a measure of intelligence, in Lipsett LP (ed): *Advances in Infancy Research*. Norwood, NJ, Ablex, 1983, vol 2, pp 31-78
4. Wilson RS: The Louisville twin study: Developmental synchronies in behavior. *Child Dev* 1983;54:293-316
5. Fagan JF: Infants' delayed recognition memory and forgetting. *J Exp Child Psychol* 1973;16:424-450
6. Terman LM, Merrill MA: *Stanford-Binet Intelligence Scale: 1973 Norms Edition*. Boston, Houghton-Mifflin, 1973
7. Dunn LM, Dunn LM: *Peabody Picture Vocabulary Test: Revised: Manual for forms L and M*. Circle Pines, MN, American Guidance Service, 1981
8. Bayley N: *The Bayley Scales of Infant Development*. New York, Psychological Corp, 1969
9. Fagan JF, Singer LT, Montie JM: An experimental selective screening device for the early detection of intellectual deficit in at-risk infants, in Frankenburg WK, Emde RN, Sullivan JW (eds): *Early Identification of Children at Risk: An International Perspective*. New York, Plenum Press, 1985, pp 257-266
10. Kopp CB, McCall RB: Stability and instability in mental performance among normal, at-risk, and handicapped infants and children, in Baltes PB, Brim OG Jr (eds): *Life-Span Development and Behavior*, New York, Academic Press, 1982, vol 4, pp 33-61
11. Ramey CT, Bryant DA, Suarez TM: Preschool compensatory education and the modification of intelligence: A critical review, in Detterman DK (ed): *Current Topics in Human Intelligence: Research Methodology*. Norwood, NJ, Ablex, 1985, pp 247-296