AN EVALUATION OF DEVELOPMENTAL EXAMINATION
AS A METHOD OF DETECTING NEUROLOGICAL,
VISUAL, AND AUDITORY HANDICAPS IN INFANCY

C. J. ROBERTS AND T. KHOSLA

Department of Social and Occupational Medicine, The Welsh National School of Medicine, Cardiff

A recent editorial (British Medical Journal, 1971) discussed the value of universal neurological screening of the newborn in the light of criteria accepted as generally applicable to all screening procedures: firstly, that a reliable screening test should not give many false negative or positive results; secondly, that early detection is worthwhile because early treatment has proved advantageous; and finally, that the use of medical and other resources is a reasonable priority in terms of total supply of these resources and the demands made on them. (It was estimated that a 15-minute neonatal neurological examination on every newborn would take up 470 hours per year of all practising paediatricians and senior paediatric registrars.)

The editorial noted that none of the requirements had been fulfilled for neonatal neurological screening but went on to say that 'It is now widely accepted that all children should have periodic developmental screening tests beginning at, say, 6 weeks, 6 months, and 10 months'. Clear and practicable schemes have been devised for doing these tests (Sheridan, 1968; Egan, Illingworth, and MacKeith, 1969).

We can find no published evidence showing how effective developmental screening (or even full developmental examination) is in detecting abnormalities. The screening schemes in current use (Frankenberg and Dodds, 1967; Sheridan, 1968; Egan et al., 1969) presume a degree of effectiveness based on the admittedly reasonable assumption that a proportion of infants with neurological, auditory, and visual defects will present as developmental delay (Gesell and Amatruda, 1941). Frankenberg and Dodds claim to have undertaken both reliability and validity studies on the Denver scale. Estimates of the former were based on the results of two tests separated by an interval of one week and undertaken by the same examiner on 20 children whose ages ranged from 2 1/2 months to 5 1/2 years (test-retest reliability) and the percentage agreement between different examiners of the same subject (examiner reliability); estimates of the validity were based on correlations between results of examination with the Denver scale and examination with the Yale developmental schedule (itself a developmental examination based on Gesell, with the additional items from various IQ tests). Important as these parameters are they clearly give no indication of the ability of developmental screening to detect clinical handicap in infancy. Assessment of reliability (or effectiveness) in this context can be determined only from concordance studies between the results of developmental screening and definitive clinical examination of the same population.

Similarly, there is no published evidence to support the criterion of advantageous early therapy in respect of infant developmental screening because of the considerable ethical and practical difficulties of undertaking controlled trials of early treatment. In this situation one must therefore be guided by the opinion of experts and accept that early treatment is of value.

Finally, it would seem plausible to argue that the reasonableness of developmental screening as a priority can only be approached when evidence of the reliability and efficacy of the procedure has been measured. Thus, in view of the current absence of evidence to support any of the above criteria for developmental screening, we felt that the one aspect which needed to be looked at urgently was that of the reliability and effectiveness of developmental examination. Such a study forms the basis of this paper.

A full developmental examination procedure is time consuming and requires considerable training and expertise for its application and interpretation. These qualifications make it unsuitable for routine use on large populations of apparently healthy children. An attempt has been made to overcome this problem by devising simplified developmental tests of items abstracted (somewhat arbitrarily) from the battery of observations in the full diagnostic
AN EVALUATION OF DEVELOPMENTAL SCREENING TESTS

95

developmental examination. It is claimed that these simplified (screening) tests are suitable for use by paramedical workers (e.g., health visitors) or even by members of the lay public who satisfy certain minimal education requirements and have undergone a short period of training (Frankenberg, 1970). Although evidence supporting the effectiveness of this approach has not yet been published and the sensitivity and specificity of the procedure are unknown, considerable resources (in terms of manpower and finance) have already been mobilised to promote this policy.

This paper is concerned with two aspects of developmental examination in relation to screening for handicaps in infancy:

(1) the extent to which full developmental examination can detect neurological, auditory, and visual defects in infants—this we have called its effectiveness (which is closely associated with reliability requirements—see above);

(2) the extent to which the number of observations made in the full developmental examination can be reduced (i.e., to make a screening test) without seriously affecting the degree of effectiveness as estimated above—this we have called redundancy.

MATERIAL AND METHODS

Data used in this study are derived from a sample originally taken to examine the influence of toxaemia and antepartum haemorrhage on infant development (Roberts, 1971). The sample consisted of 193 infants aged 11 to 13 months whose mothers experienced toxaemia and/or antepartum haemorrhage during the related pregnancy, and 193 controls matched for sex, parity, maternal age, place of birth, and social class whose mothers did not experience toxaemia or antepartum haemorrhage.

One of us (C.J.R.) carried out a four-part examination on each subject:

(1) a developmental examination using the Gesell developmental test (Gesell and Amatruda, 1941), which consists of items divided into five subgroups: gross motor ($x_1$), fine motor ($x_2$); adaptive ($x_3$); language ($x_4$); and personal social ($x_5$). In each subject the developmental age score is the ratio of the number of items completed successfully compared with those required by Gesell at the same age. The aggregate developmental score is based on the mean of the five subgroups:

$$y = \frac{1}{5} (x_1 + x_2 + x_3 + x_4 + x_5);$$

(2) a neurological examination using a battery of 64 observations covering muscle tone, reflex activity, and neuromotor patterning (Gesell and Amatruda, 1941; Knobloch and Pasamanick, 1962);

(3) an examination of auditory responses to free field pure tones (Roberts, 1968). A history of previous ill health (to include upper respiratory tract and ear infections) was taken. Otoscopy was done on all subjects whose auditory responses were abnormal;

(4) an examination of binocular visual competence using a ½in diameter white ball rolling horizontally across the infant’s line of vision (Sheridan, 1960) together with an examination of the eyes for nystagmus, failure of fixation, use of peripheral vision, and strabismus.

Data thus collected were analysed:

(1) for evidence of redundant tests in the full developmental examination (by use of a multiple regression technique); this analysis was undertaken on the 193 control subjects in the original sample (subsample 1);

(2) to consider the degree of concordance between the clinical classification of abnormality based on independent neurological criteria and the developmental classification (in percentiles) based on the full developmental examination. This analysis was undertaken on all infants from the original sample who were classified as abnormal on the basis of the neurological examination (the group comprised 12 infants and will be referred to as subsample 2);

(3) to study the concordance between the results of auditory assessment and language development (subsample 3);

(4) to study the concordance between visual examination and fine motor development (subsample 4). Both subsamples 3 and 4 included every infant in the original sample (i.e., cases and controls, $n = 386$).

Support for the inclusion of all subgroups ($x_1$ to $x_5$) in both the full and the simplified (screening) developmental examination derives from the belief that while some children (e.g., those with mental handicap) will show delay in all subgroups, others (e.g., those with motor, visual, or hearing handicap) will appear abnormal on one subgroup only; but it could equally be argued that a defect in one field large enough to be detected by these relatively crude techniques is likely to have an effect in other fields as well.
RESULTS

MULTIPLE REGRESSION ANALYSIS (SUBSAMPLE 1)

Table I shows the subgroup and aggregate developmental scores for subsample 1, and Table II shows the intercorrelation between these. The size of some of these correlations (viz., \( x_4 \) and \( x_5 \); \( x_2 \) and \( x_6 \)) suggests that certain subgroups could be deleted from the full developmental examination without any appreciable loss of information.

Table III considers how much information is lost from the aggregate developmental score as various subgroups are omitted using the method of stepwise regression. The square of the multiple correlation coefficient \( R^2 \) accounts for the percentage variation in ‘\( y \)’ (the aggregate score) accounted for by the set of ‘independent’ variables. Thus, if all the five subgroups are included in the examination, \( R^2 \) should be 100%—the observed \( R^2 \) (98-12) is attributable to errors due to rounding.

Deletion of \( x_4 \) and \( x_5 \) shows a negligible loss of information; deletion of \( x_4 \) results in a 10% loss of information in ‘\( y \)’; and deletion of \( x_5 \) causes a further 20% loss of information in ‘\( y \)’. Subgroups \( x_4 \) and \( x_5 \) together give more information than the four subgroups \( x_2, x_3, x_4, x_5 \); \( x_4, x_5; x_1, x_3, x_5; x_1, x_4, x_5 \) but all yielded appreciably less information than \( x_1, x_5, x_4 \). Thus, three developmental subgroups, \( x_1, x_5, x_4 \), and \( x_2 \), gave as much information as all five subgroups, suggesting that tests \( x_2 \) and \( x_4 \) may be superfluous in the routine examination of the year-old infant.

ANALYSIS OF DATA FROM SUBSAMPLE 2

The results described so far refer to the routine examination of predominantly normal children (only two subjects in subsample 1 had a neurological defect). It could be argued that subgroups considered redundant (\( x_5 \) and \( x_6 \)) or of minimal usefulness (\( x_4 \)) in respect of normal children may well be of value for screening abnormal children. In order to study this, the concordance between neurological classification and developmental classification was examined. The scores of the 12 subjects in subsample 2 (A to L) in each of the five developmental subgroups are shown in Table IV.

Table V shows the distribution of subsample 2 by their percentiles. To illustrate, subject C scored 60 on \( x_1 \) (Table IV) and thus fell below the first percentile. Seven of the 12 subjects were classified abnormal (i.e., below the 20th percentile) by \( x_5 \); five by \( x_4 \); seven by \( x_5 \); in certain of these the classification was borderline (i.e., between the 10th and
AN EVALUATION OF DEVELOPMENTAL SCREENING TESTS

TABLE V
DEVELOPMENTAL SCORES IN PERCENTILES OF 12 INFANTS CLASSIFIED AS ABNORMAL BY INDEPENDENT NEUROLOGICAL CRITERIA

<table>
<thead>
<tr>
<th>Subject</th>
<th>Developmental Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x₁</td>
</tr>
<tr>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>10</td>
</tr>
<tr>
<td>I</td>
<td>1</td>
</tr>
<tr>
<td>J</td>
<td>10</td>
</tr>
<tr>
<td>K</td>
<td>1</td>
</tr>
<tr>
<td>L</td>
<td>20</td>
</tr>
</tbody>
</table>

Percentile code:
- 1 = below 1st percentile
- 5 = below 5th percentile
- 10 = below 10th percentile
- 20 = below 20th percentile
- N = above 20th percentile

20th percentile—three in x₁, one in x₂, and two in x₃. In every case where subjects were classified abnormal by x₃, x₄, and x₅ they were also classified abnormal by x₁. Two (B and I) who were not abnormal in x₁, x₄, and x₅ were abnormal in x₁; and two (C and L) were normal in x₄ but abnormal in x₁.

Table V shows that three infants had abnormal results in x₁, one in x₂, and two in x₃. In every case where subjects were classified abnormal by x₃, x₄, and x₅ they were also classified abnormal by x₁. Two (B and I) who were not abnormal in x₁, x₄, and x₅ were abnormal in x₁; and two (C and L) were normal in x₄ but abnormal in x₁.

Subject A was classified normal by developmental examination but abnormal by neurological examination; he was found to have abnormal neuromotor patterns confined to the legs which were not severe enough to delay motor development below the 20th percentile.

Thus the 12 infants classified as clinically abnormal on the basis of independent neurological criteria were classified as follows on the basis of the developmental examination:

Test | Classification based on Developmental Percentiles |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abnormal (below 10th)</td>
</tr>
<tr>
<td>x₁</td>
<td>10</td>
</tr>
<tr>
<td>x₂</td>
<td>7</td>
</tr>
<tr>
<td>x₃</td>
<td>5</td>
</tr>
<tr>
<td>x₄</td>
<td>4</td>
</tr>
<tr>
<td>x₅</td>
<td>4</td>
</tr>
</tbody>
</table>

These results suggest that test x₁ is of overwhelming importance in screening for neurological abnormality by developmental examination. Tests x₁, x₃, and x₅ are not specific for neurological abnormality and thus tend to misclassify many infants as normal or of only borderline significance. Test x₄ did not provide any information that was not already available from test x₁.

ANALYSIS OF DATA FROM SUBSAMPLE 3
Twenty-one subjects had abnormal responses to sound. Of these
(a) eleven (28 7% of total sample) were considered to have conductive deafness due to middle ear disease, a diagnosis based on the triad of abnormal hearing test result, otoscopic evidence, and a past history of middle ear disease;
(b) four (1 0%) had high frequency deafness diagnosed on the basis of consistent differential responses to high (2 and 4 kHz) compared with low and middle (0 25, 0 5, and 1 kHz) pure tones;
(c) four (1 0%) had abnormal responses not considered to be caused by hearing loss but by general developmental or specific motor retardation.

Thus in this study 5 4% of the sample failed the definitive hearing test, of whom one quarter were considered to be false positives (group c above). No cases of severe congenital deafness were encountered (expected frequency 0 01% Fraser 1964).

The two remaining subjects with abnormal responses to sound showed no distinctive clinical, neurological, or developmental features.

The concordance between the results of definitive auditory assessment on the one hand and language and motor development on the other is shown in Table VI. Fifty-seven per cent of subjects who failed

TABLE VI
DISTRIBUTION OF 21 INFANTS WITH ABNORMAL RESPONSES TO SOUND BY RESULTS ON TEST x₁ (MOTOR DEVELOPMENT) AND TEST x₄ (LANGUAGE DEVELOPMENT)

<table>
<thead>
<tr>
<th>Test x₁</th>
<th>Test x₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below 20th Percentile</td>
<td>Above 20th Percentile</td>
</tr>
<tr>
<td>Below 20th percentile</td>
<td>4</td>
</tr>
<tr>
<td>Above 20th percentile</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
</tr>
</tbody>
</table>

the hearing test had normal language development (i.e., above the 20th percentile); four out of the remaining nine subjects with delayed language development also had delayed motor development (all category c above) and would, therefore, have been screened as abnormal using test x₁ above.

ANALYSIS OF DATA FROM SUBSAMPLE 4
Three hundred and eighty-three subjects (99 2%) had normal visual competence; of the three subjects who failed the rolling ball test, one could not sit upright at all and two had severe head drop when supported in the sitting position; but when tested in the supine position with a 1 in diameter dangling
ring moved across the line of vision, all three followed the test object which suggested that their visual competence was normal for their level of motor development. No gross abnormalities of the eyes (i.e., nystagmus, cataract, limitation of peripheral vision, or ocular malformation) were observed in any subject but 10 (2.6%) had strabismus. None of those with strabismus had abnormal visual competence; their mean fine motor developmental score (test \( x_2 \)) was 96.4 compared with 95.5 for the remainder of the sample. The difference is not statistically significant.

In this study there was no concordance between visual examination (i.e., inspection of the eyes for cataract, nystagmus, squint, etc.) and fine motor (coordinative) development. (It is accepted that gross visual defects are likely to delay development (Gardiner, MacKeith, and Smith, 1969) but this could not be studied in this project since no such gross defects were observed.)

**Discussion**

Estimates of effectiveness can be made by comparing the results of the exposure of a population first to the developmental testing procedure and then to an examination using clinical criteria of neurological, auditory, and visual abnormality. So far as we can determine, neither the full developmental examination procedures (Gesell and Amatruda, 1941; Griffiths, 1954) nor currently used developmental screening tests consisting of items abstracted from the battery of observations listed in the full developmental examination (Sheridan, 1968; Egan et al., 1969; Frankenberg, 1970) have yet been validated in this way, and consequently their true value remains in some doubt. In particular, the notion that the screening procedure should include observations from each of the subgroups in the full developmental examination in order to detect subjects with a single motor, visual or hearing handicap has never been proven experimentally.

**Effectiveness of Developmental Examination**

In the present study this has been estimated by observing the number of children with neurological, auditory, and visual defects discovered by clinical examination who would also have been classified as abnormal on the basis of developmental tests alone:

1. Ninety-two per cent of infants with clinical neurological defects were classified as abnormal on the basis of developmental tests alone, these being subjects who showed retardation in gross and fine motor subgroups \( (x_1 \) and \( x_2 \)). This is perhaps not altogether surprising since, clinically, abnormal signs are essentially observations of motor function and thus developmental examination is likely to prove to be a highly effective method of eliciting abnormal neurological signs in infancy (Gesell and Amatruda, 1941; Knobloch and Pasamanick, 1962).

2. Only 43% of subjects with independently determined hearing loss (i.e., by definitive free-field hearing tests) were classified as abnormal on the basis of a developmental examination, i.e., language development (test \( x_4 \)). Possible explanations for normal language development in the presence of abnormal hearing are—deafness not severe enough to delay early linguistic development (at 1 year); deafness of recent origin; and deafness of an intermittent nature (e.g., due to middle ear effusion). However, hearing disorders such as these may well go on to cause speech, behaviour, and educational problems, and there is little doubt that these defects (rather than the more severe and much rarer forms of congenital neurosensory deafness) form the bulk of the work of the school and peripatetic services for the deaf (Robinson et al., 1967). The results of this study, therefore, suggest that developmental testing is not an effective method of detecting this important category of hearing disorders in infancy.

3. No subjects with visual defects were recorded as having abnormal development. No cases of severe defects were observed in this study, but if, as Gardiner et al. (1969) suggest, gross disorders of eyesight manifest themselves in infancy in terms of slow development, it can be reasonably assumed that it would be in errors of fine and gross motor function (tests \( x_1 \) and \( x_2 \)). In a paper on visual screening in very young children, Sheridan (1960) stated that there was general agreement (presumably among paediatricians) that the earlier all visual defects were diagnosed and treated the more favourable is their prognosis; more recently, Gardner et al. (1969) referred to children with squints, amblyopia, and myopia as neglected ‘because tests of vision are generally omitted from screening procedures in infants and young children’. Accepting the view of these authorities that early diagnosis of all visual defects is of value since it will lead to a substantial improvement in prognosis, the results of this study suggest that developmental testing is not an effective method of detecting this category of defects.

Thus, in summarizing the effectiveness of this form of examination it would seem that developmental testing at 1 year is effective in detecting neurological defects but not in discovering abnormalities of vision and hearing.
REDUNDANCY

Having determined which categories of defect the developmental examination can effectively detect, it remains to determine the smallest number of observations required before the degree of effectiveness is significantly reduced.

The results of analysis of subsample 1 show that two time-scaled behavioural observations (i.e., gross and fine motor function) would predict 87% of the total variance of the full developmental score calculated as an aggregate of the five subgroup scores, and would also detect 92% of infants classified abnormal by independent neurological examination. One reason for this is that many test items merely repeat, or are minimal variations of, observations made in other subgroups with the result that the information they provide is not additive.

Joint consideration of effectiveness and redundancy of developmental examination suggests that the information effectively provided by full developmental examination could be equally well provided by a much simplified developmental test consisting of observations of gross and fine motor function (see above). This procedure may have certain advantages—it utilizes observations which are easily defined and do not depend on maternal recall and so are thus likely to have low observer error. For these reasons it may well be suited for use by paramedical or even lay personnel. Also the procedure time would be reduced to approximately one minute compared with 20 to 30 minutes for existing developmental screening tests, thus allowing more time to be spent (if considered necessary) on specific tests of auditory and visual function.

The number and complexity of the observations which make up a screening procedure will determine the number and professional status of the staff required. This will influence the cost of the procedure and this in turn will determine whether it is adopted nationally or locally, and, if adopted, whether it reaches all those who may benefit from it or only certain selected groups (e.g., at risk categories). At present most developmental testing (and hence screening) procedures are rather complicated and time consuming and thus usually have to be applied by medical or paramedical personnel with the result that they reach only a small proportion of the infant population. In view of the apparent specific range of effectiveness of developmental testing (and hence developmental screening) in the detection of handicaps in infancy (i.e., developmental and neurological but not auditory or visual), there may have been an over investment (in terms of the number and professional status of personnel employed, and the time spent in testing individual subjects) on this aspect of paediatric screening. The evidence collected in this study suggests that time spent in developmental screening might be usefully reallocated—much less to be spent on the examination of development (which could be accomplished with a simple screening test such as that proposed in this paper undertaken by trained but non-medical observers) and more on a detailed clinical examination of auditory and visual function.

Some handicaps of infancy are uncommon (e.g., congenital neurosensory deafness and severe visual defects) and it might be reasonably argued that in view of the relatively small size of the sample studied in the project it cannot necessarily be concluded that developmental screening is ineffective in detecting such disorders. We accept this criticism but believe that, in view of the considerable and growing deployment of medical resources into developmental screening, the onus lies with the proponents of this approach to show by concordance studies (on necessarily large populations) that the recommended procedures are effective and reliable.

SUMMARY

The simplicity (measured in terms of number and nature of observations required, procedure time and professional status of tester) of developmental screening tests will largely determine the extent to which they can be applied to every member of the infant community. In the search for simplicity the present study looks at two aspects of developmental examination in relation to handicap at 1 year. One is the extent of redundancy among component observations of the developmental test (i.e., can the same amount of information be obtained from fewer observations). The other is the effectiveness of developmental testing estimated from the concordance between the results of full developmental examination and examination using clinical criteria of neurological, visual, and auditory abnormality.

The results of the analysis, based on a sample of 386 infants aged 11 to 13 months, show that developmental examination has only limited effectiveness when applied during infancy (i.e., in the study 92% of all subjects with neurological defect, but only 40% of subjects with abnormal hearing, and no infants with visual defects were classified as abnormal on the basis of a full developmental examination). Furthermore, there was clear evidence of considerable redundancy among component observations of the full developmental examination—a multiple regression analysis showed that 87% of the total variance of the aggregate developmental
score derived from the five components of the full examination could be predicted from two component observations (gross and fine motor function) without reducing the levels of effectiveness for the individual categories of handicap quoted above. The findings of the small study reported in this paper lead us to conclude that medical and para-medical personnel currently engaged in the routine application of developmental tests during infancy may not be effectively deployed (and therefore in relation to the scarcity of health service manpower may possibly be inefficiently deployed) since they spend considerable time making many observations which contribute nothing to the case-finding effectiveness of the procedure, and the small number of observations that are effective are of such a nature that they could probably be made equally well by lay personnel. In view of the importance of establishing that the use of medical and para-medical resources in infant screening is a reasonable priority (in terms of total supply of resources and demands made upon them), there would appear to be an urgent need for a large-scale study of the effectiveness of developmental screening as a method of detecting handicaps in infancy.

The research was sponsored by a generous grant from the Association for the Aid of Crippled Children, New York.

We are indebted to Dr. W. E. Thomas, Medical Officer of Health for Glamorgan, for allowing us both access to the medical records and the opportunity to liaise with the health visitor staff working in the area where the research was undertaken.

REFERENCES


