The Denver Developmental Screening Test Compared With the Stanford-Binet Test

JOHN A. GRANT, M.D., M.P.H., and ALAN M. GITTELSOHN, Ph.D.

SYSTEMATIC screening of patient populations to facilitate clinical services had its practical beginning with the Kaiser-Permanente Clinic (*1*) in the early 1950's. This approach had not been seriously applied to pediatric age groups, however, until Frankenburg and Dodds (*2*) introduced the Denver Developmental Screening Test (DDST) in 1967. In recent years, this test has become widely used by a variety of professional and nonprofessional workers in pediatric clinics, at maternal and child health conferences, and in children and youth projects.

Dr. Grant is chief physician and deputy director, Community Pediatric Center, University of Maryland School of Medicine, and former chief, school health section, division of maternal and child health, Maryland State Department of Health. Dr. Gittelsohn is professor of biostatistics, Johns Hopkins University School of Hygiene and Public Health. Tearsheet requests to John A. Grant, M.D., Department of Pediatrics, University of Maryland School of Medicine (attn: Mrs. Mantik), Redwood and Greene Sts., Baltimore, Md. 21201.

In their early publication, Frankenburg and Dodds compared the performance of the Denver Developmental Screening Test in a study population of 18 children to the Revised Yale Developmental Schedule (*3*). They found high correlations but did not examine the performance of the test in more explicit administrative terms of pathology discovered or pathology missed.

We show how the performance of the Denver Developmental Screening Test was evaluated for a specific population and relate this kind of evaluation to the needs of the service program administrator.

Method and Materials

The test was given to 590 children, 3½ to 4½ years of age, during an epidemiologic study of congenital rubella (the parent study) in the summer of 1968 (*4*). The children were chosen to represent the population of a semirural Maryland county; they came from a wide variety of socioeconomic strata, as measured by a two-index (education and occupation) Hollingshead scale (table 1).

The primary goal of the parent study was to determine the prevalence of congenital rubella in the sample population by means of a five-part screening battery. The Denver Developmental Screening Test was one part of the battery. The test was administered, under close supervision, by four specially trained technicians and scored according to the manual provided by Frankenburg and Dodds in 1968. Toward the end of the parent study the staff became concerned about the field performance of the test because of an apparently high overreferral rate, which raised parallel concern about underreferral—a more serious error in view of the primary study goal. We therefore decided to evaluate the performance of this screening test in attempting to correctly identify the intelligence of a typical group of preschool children as measured by a standard test.

Planning, funding, and relocating subjects from the parent study consumed 12 months—longer than anticipated. Samples of 160 DDST "normals," 75 DDST "questionables," and 63 DDST "abnormals" were located and given the Stanford-Binet Test of Intelligence by a team of four experienced psychologists, on leave from the Baltimore City schools, under the direction of Lewis Armistead, Ph.D., Maryland State Department of Health. The distribution of 298 IQ scores is detailed in table 2. It is slightly skewed toward the higher score intervals, reflecting the distribution of socioeconomic status in the study population.

Since the data needed to calculate the Hollingshead scale (table 1) were not available for either the county under study or for the State as a whole, the percentage distribution of Hollingshead's original study group is given as a frame of reference. As would be expected in comparing an urban industrial community with a semirural county, a shift is seen in the distribution of the semirural county from class IV to class II, thus giving reassurance that a representative sample was being studied.

Table 2. Distribution of 298 actual IQ scores in this study as compared with expected normal distribution

Stanford-Binet IQ scores	Expected percent distribution	Actual percent distribution
0–67	2.2	0.7
68–83	13.6	7.4
84–99	34.2	37.9
100–115	34.2	35.2
116–131	13.6	16.4
132+	2.2	2.4
Total	100.0	100.0

DDST normals, abnormals, and questionables were defined, according to Frankenburg and Dodds' manual, as follows:

DDST normal: no items failed below the child's age level; at least one item, dissected by the age line, passed.

DDST abnormal: at least two items failed below the child's age level.

DDST questionable: all others.

A child's total DDST rating was the lowest rating in any one of the four sections: gross motor, fine motor, language, or personal-social. Tasks that were refused were considered questionable.

Table 1. Range and distribution of socioeconomic status of 590 study subjects, as measured by two-index (education and occupation) Hollingshead scale

Social class	Score range, this study	Percent distribution, this study	Percent distribution, Hollingshead's New Haven community [1]
I, professional and management	11–13	3.1	2.7
II, business and semiprofessional	14–30	14.8	9.8
III, clerical and sales	31–49	21.9	18.9
IV, skilled workers	50–64	37.8	48.4
V, unskilled workers and unemployed	65–77	22.4	20.2
Total		100.0	100.0

[1] See text above.
SOURCE: Reference 5.

To analyze the ability of the Denver Developmental Screening Test to predict intelligence as measured by the Stanford-Binet test, we dichotomized the results of both reference and screening tests as follows:

Stanford-Binet Test of Intelligence:
- Positive = 0 to 83
- Negative $\geq$ 84

Denver Developmental Screening Test:
- Positive = questionable and abnormal
- Negative = normal

Table 3 shows the screening test-reference test two-by-two contingency table, from which various evaluative indices may be calculated (6, 7). The indices of sensitivity, specificity, underreferral, overreferral, and tetrachoric r (cosine-pi approximation) selected for this study were calculated as follows: Sensitivity $= \frac{a}{a+c}$, specificity $= \frac{d}{b+d}$, underreferal $= \frac{c}{c+d}$, overreferral $= \frac{b}{a+b}$, and tetrachoric r (cosine-pi approximation) =

$$\cos\left(\frac{180°}{1+\sqrt{\frac{ad}{bc}}}\right).$$

Section a of table 3 indicates true positives, section b false positives, section c false negatives, and section d true negatives. Sensitivity indicates the ability of a screening test to detect a true positive result out of all diseased persons. Specificity indicates the ability of a screening test to detect a true negative result out of all nondiseased persons.

Underreferral is the ratio of persons screened negative with disease to total persons not re-

Table 3. Two-by-two contingency table from which various evaluative indices may be calculated

Screening test	Reference test		Total
	Positive	Negative	
Positive	a	b	$a+b$
Negative	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

NOTE: See text for calculations.

Table 4. Study data used for the two-by-two contingency table

Denver Developmental Screening Test	Stanford-Binet test		Total
	Positive	Negative	
Positive	26	112	138
Negative	7	153	160
Total	33	265	298

ferred. Overreferral is the ratio of persons screened positive with no disease to total persons referred.

Tetrachoric r is a correlation coefficient that provides a convenient method of estimating r from data that are in the form of continuous measurements, as used in this study. A perfect positive correlation between screening test and reference test would be indicated by a value of +1.0.

Results

The study data used for the two-by-two contingency table are shown in table 4. Calculation of the selected evaluative indices revealed a sensitivity of .788, a specificity of .578, an underreferral ratio of .044, an overreferral ratio of .812, and a tetrachoric r of .581.

A different breakdown of the data is shown in table 5, where each of the four sections of the test are compared with the Stanford-Binet results as if they were the complete test. This was done to examine the relative contribution of various combinations of the individual sections of the test to correlation with the Stanford-Binet test.

Discussion

This evaluation of the Denver Developmental Screening Test was an outgrowth of a larger epidemiologic study of educational casualty resulting from congenital rubella (*4*) in which children were selected for formal intelligence testing on the basis of DDST performance. It was essential, therefore, that the validity of the screening instrument be determined under actual field conditions. Since these circumstances were similar to those that would prevail, for example, in a county health department, we felt that our findings would be of value to public health administrators interested in child development.

The sensitivity of the Denver Developmental Screening Test in this study indicated that it performed well under field conditions, correctly identifying eight of 10 children with low IQ's. Specificity was somewhat low, however, with a high overreferral ratio anticipated. The underreferral ratio was reassuringly low, especially in view of the primary goal of the parent study. The value of tetrachoric r (.581) calculated from these data compared favorably (.590) with the well-known Massachusetts Vision Test (*7*), giving the Denver Developmental Screening Test a good overall rating for program use.

Comparison of the evaluative profiles of various combinations of the sections of the Denver Developmental Screening Test (table 5) produced several findings that deserve comment. The gross motor and personal-social sec-

Table 5. Profile comparisons of various combinations of sections of the Denver Developmental Screening Test

Evaluation	Complete test	Fine motor and language	Gross motor	Fine motor	Language	Personal-social
Sensitivity	.788	.708	.208	.542	.417	.125
Specificity	.587	.697	.812	.771	.827	.875
Underreferral	.044	.036	.079	.050	.059	.081
Overreferral	.812	.828	.911	.827	.825	.919
Tetrachoric r	.581	.604	.084	.500	.482	.000

tions contribute little or nothing to correlation of the test with the Stanford-Binet IQ test. The fine motor and language sections individually show higher correlation, but a combination of these two sections is superior to the test as a whole. This result is not surprising because the Stanford-Binet test relies heavily on fine motor and language functions.

Two important limitations of this evaluation should be noted: (*a*) the screening and the reference tests were separated by a period of 12 months and (*b*) some DDST tasks (for example, gross motor and personal-social) were not designed to be predictors of intelligence quotients, as measured by the Stanford-Binet test. Even though the screening tests are age specific, the variance would be expected to be greater in a given group of children at an earlier age. Also, some mothers may have schooled their children in several of the fine motor and language tasks of the test that are derived from the Stanford-Binet test.

These limitations, in addition to low *r* values in the gross motor and personal-social sections would help to explain the relative superiority of combining the fine motor and language sections as a predictor of a Stanford-Binet intelligence score. Further, to obviate the fact that we have related this screening test to only one measure of intelligence (the Stanford-Binet test), a followup study of this cohort to examine other parameters of social, emotional, educational, and neurological function is planned for 1972–73.

REFERENCES

(*1*) Collen, M. F., and Linden, C.: Screening in a group practice prepaid medical care plan. J Chronic Dis 2: 400–408, October 1955.

(*2*) Frankenburg, W. F., and Dodds, J. B.: The Denver Developmental Screening Test. J Pediatr 71: 181–191, August 1967.

(*3*) Gesell, A., and Amatruda, C. S.: Developmental diagnosis. Ed. 2. Harper & Row, New York, 1958.

(*4*) Grant, J. A., and Gittelsohn, A. M.: Estimating educational casualty from the 1964 rubella epidemic in Maryland. Md State Med J. In press.

(*5*) Hollingshead, A. B., and Redlich, F. C.: Social class and mental illness. John Wiley & Sons, Inc., New York, 1958.

(*6*) Thorner, R. M., and Remein, Q. R.: Principles and procedures in the evaluation of screening for disease. PHS Publication No. 846, Public Health Monogr No. 67. U.S. Government Printing Office, Washington, D.C., 1961.

(*7*) Blum, H. L., Peters, H. B., and Bettman, J. W.: The Orinda Study: Vision screening for elementary schools. University of California Press, Los Angeles, 1959.

GRANT, JOHN A. (University of Maryland School of Medicine), and GITTELSOHN, ALAN M.:

The Denver Developmental Screening Test compared with the Stanford-Binet test: Evaluation of 298 preschool children. Health Services Reports, Vol. 87, May 1972, pp 473–476.

The Denver Developmental Screening Test (DDST) was evaluated by comparing the DDST performance of 298 children with their performance under the Stanford-Binet Test of Intelligence. The classic epidemiologic profile, including sensitivity (.788), specificity (.578), underreferral (.044), overreferral (.812) and tetrachoric *r* (.581) was used.

Under these circumstances (testing of a discreet population at a specific point in time) the Denver Developmental Screening Test was shown to be a useful procedure for identifying preschool children with low Stanford-Binet intelligence scores, but it had a strong tendency to overrefer.

Limitations of this evaluation were a 12-month interval between screening and reference tests and the use of a broad spectrum test like the Denver Developmental Screening Test to predict a single parameter of development—intelligence as measured by the Stanford-Binet test. The evaluation does, however, reflect performance of the test under actual field conditions, providing the administrator with valuable information for program control.