BRIGANCE®

Technical Report for the BRIGANCE® Screens

Frances Page Glascoe, Ph.D.
Vanderbilt University
About the Author

Dr. Frances Page Glascoe is a professor of pediatrics at Vanderbilt University and a pediatric educator with more than twenty years experience in working with pediatric residents, faculty, and community providers. She spent ten years teaching special education in the public schools before joining the faculty at Vanderbilt University where she directed the rotation in developmental pediatrics and served as educational specialist on the neuropsychology, school performance, autism, and developmental diagnostic teams. Her research focuses on the accuracy of developmental and behavioral screening measures and she is the author of more than 100 journal articles and chapters.

Dr. Glascoe is the author or co-author of several screening tests including Parents’ Evaluation of Developmental Status (PEDS) and PEDS: Developmental Milestones, both brief screening, surveillance and decision-support tools focused on detecting and addressing developmental and behavioral needs in children 0–8 years (www.pedstest.com), the BRIGANCE® Screens, and the Safety Word Inventory and Literacy Screener (SWILS). In October 2000, she received the Dale Richmond Award for contributions in child development from the American Academy of Pediatrics.

Acknowledgments

Thanks to the many school systems and personnel who assisted in the revisions and validations of the BRIGANCE® Screens. They are listed by name and location in each of the test books.
CHAPTER 2. General Administration Procedures for the BRIGANCE® Screens

Screening and Evaluation Process

Figure 2-1 shows the general process of data collection using the BRIGANCE® Screens. The diagram illustrates the sequence and types of information to be collected and the process of making referrals.

Figure 2-1. Screening and Evaluation Flowchart

DATA COLLECTION

- Administration of age-appropriate BRIGANCE Screen (REQUIRED)
- Registration and Information Form to obtain broad data on each child
- Supplemental Assessments* to assess higher-level skills when needed
- Screening Observations Form*
- Teacher’s Rating Form*
- Parent’s Rating Form*
- Data from nurse, physician, speech therapist, etc.
- Self-help and Social-Emotional Scales and Reading Readiness Scale*

*optional

REVIEW/EVALUATION

- Screening team review and evaluation of data

PLACEMENT/REFERRAL

- Placement
- Referral
CHAPTER 7. Standardization of the BRIGANCE® Screens

Critical Concepts in Screening and Test Construction: Standardization

- The directions for scoring and administration have been field tested and are clear enough that the test can be administered in exactly the same way by different examiners.
- The test has been administered to many children, who represent the geographic regions of the United States and the demographic characteristics of the U.S. population as a whole.
- The parents of children in the standardization sample represent the current U.S. population in terms of educational attainment, ethnicity, income levels, and other demographic characteristics.
- Results of testing reflect the average performance of children according to ethnicity, gender, age, socioeconomic differences, etc.

History

In 1979, Albert Brigance selected items from the BRIGANCE® Diagnostic Inventory of Early Development (IED) to create the BRIGANCE Screens. The IED is a lengthy battery of tasks tapping language, motor, cognitive, self-help, academic, and readiness skills and is widely used in curriculum planning and in the development of IEPs. Because BRIGANCE Screen items were selected from the IED, its standardization forms the basis for the psychometrics of the BRIGANCE Screens.

The original IED was standardized on 1,156 children ranging in age from one year, one month to beyond six years. The group was stratified by gender (50% male), race (73% White, 15% African-American, and 12% Hispanic), urban/rural/suburban areas, and geographic location (Northeast, Southeast, and Western U.S. states). A technical report prepared by Dr. Brian Enright at the University of North Carolina at Charlotte (available from Curriculum Associates, Inc.) presents the mean performance of children on each section and task of the IED.

The IED items selected for the BRIGANCE Screens were rated by a large group of teachers, diagnosticians, and curriculum supervisors across the United States according to the degree of correspondence between items and curriculum objectives. Items were retained if they were nominated by at least 90% of the professionals. Field testing was then conducted in thirty-five different schools or districts and thirteen states spanning the geographic regions of the United States, using an experimental version of the screens. The results of the field trials were used to finalize item selection, clarify item content, and refine directions for administration and scoring.
1995 Restandardization of the *BRIGANCE®* Screens and 2001 Standardization of the Infant and Toddler Basic Assessments

Due to the need for updated standardization and validation data, Curriculum Associates, Inc. commissioned research on the *BRIGANCE®* Screens. Completed in 1995, the study was designed to define the range of children's performance across geographical boundaries, socioeconomic levels, and other variables. In 1998, work began on the development, standardization, and validation of the *BRIGANCE* Infant and Toddler Basic Assessments and was completed in 2001. The process used to collect data is described below.

**RECRUITMENT.** In the 1995 standardization study, four sites were selected to represent the broad geographic regions of the United States: North (Plymouth, Massachusetts); Central (Denver, Colorado); South (Tampa, Florida); and West (Carson City, Nevada). At each site, 114 children and their families were recruited, although actual participation varied slightly across sites (generally because of personnel shortages) to produce a total of 408 subjects. Within each site, one to two schools were identified that had a balance of children from high, middle, and low socioeconomic status (approximately one-third of students participated in the federal free/reduced lunch program). At each school, one kindergarten and one first-grade classroom were identified and students with parental consent were enrolled in the study. Children, aged two years through four years, were recruited if they were the siblings of children in the target schools. However, in two sites an insufficient number of younger siblings were available, and so recruitment was extended to children attending preschool programs in the zone of the targeted elementary schools. These programs were also selected because they: (a) had federal or local funding subsidies; (b) served children from varying socioeconomic backgrounds; and (c) were neither oriented for special education students nor exclusive of children with known disabilities.

In the 2001 standardization study, twenty-nine sites were recruited to sample the performance of children in the birth to two-year range on the newly created Infant and Toddler Basic Assessments. Sites included pediatric offices, public health clinics, child-find programs, day care and preschool programs, and university-affiliated assessment clinics in twenty-one states. At each site, examiners were asked to select consecutive children and families, obtain informed consent, and administer the Infant and Toddler Basic Assessments to each.
CHAPTER 8. Reliability of the BRIGANCE® Screens

Critical Concepts in Screening and Test Construction: Reliability

- **Internal consistency** tells how well items on each of the BRIGANCE® Screens relate to the total test score, suggesting that the Screens are measuring a well-defined construct (i.e., developmental and readiness skills).

- **Test-retest reliability** answers the question, If you give the BRIGANCE Screens several days to several weeks later to the same child, do you get roughly the same score?

- **Stability** provides information on the relationship between results when a test is readministered several months apart.

- **Inter-rater or inter-examiner reliability** tells whether roughly the same score is obtained if two different examiners retest the same child.

**Internal Consistency**

**2005 STANDARDIZATION AND VALIDATION STUDIES.**

For the current study, Guttman scalability coefficients were produced for total scores on the Basic Assessments in each of the Screens. The coefficients serve as an indicator that the Basic Assessments and their items are hierarchical, unidimensional, homogeneous measures of academic and readiness skills. The standard errors of measurement (SEₘ) provide a confidence band around the derived score that accounts for fluctuations in reliability due to measurement error. SEₘs are both added to and subtracted from each student's total score in order to provide a theoretical, error-free indicator of true performance. Coefficients are influenced by test length which explains the lower lambda values for shorter scales.

**Table 8-1. Internal Consistency and Standard Error of Measurement of Total Scores**

<table>
<thead>
<tr>
<th>Basic Assessments</th>
<th>Infant</th>
<th>Toddler</th>
<th>Two-Year-Old Child</th>
<th>Two-and-a-Half Year-Old Child</th>
<th>Three-Year-Old Child</th>
<th>Four-Year-Old Child</th>
<th>Five-Year-Old Child/Kindergarten</th>
<th>First Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guttman Lambda Coefficient</td>
<td>.97</td>
<td>.94</td>
<td>.98</td>
<td>.97</td>
<td>.99</td>
<td>.99</td>
<td>.90</td>
<td>.99</td>
</tr>
<tr>
<td>Standard Error of Measurement</td>
<td>0.86</td>
<td>1.01</td>
<td>2.70</td>
<td>1.43</td>
<td>1.29</td>
<td>1.26</td>
<td>1.60</td>
<td>1.70</td>
</tr>
</tbody>
</table>
The standard error of measurement (SEM) accounts for fluctuations in reliability due to measurement error and shows how much a person’s observed score is likely to differ from his true score, particularly if the same test were administered multiple times. The standard error of measurement is used to build a confidence interval around the observed score. This enables a range to be computed that will include the true or actual score with a specified level of confidence.

For example, if a child has a score of 18 on an assessment with a standard error of measurement of 2, the standard error of measurement of 2 is used to place a confidence band of \(18 \pm 2\) around the observed score. This is the 68% confidence level and thus it can be said that the true score would be in that range (16–20). Or we can say with 95% confidence that the student’s true score lies in an interval within two standard errors of measurement of the observed score (18 ± 4 = 14–22). Put another way, 95 times out of 100, the re-test score would be between 14 and 22.

The SEM itself, when banding the observed score, functions as a confidence level of 68%. It is also possible to band the observed score with higher confidence levels, e.g., 90% or 95% as described in Chapter 3. All confidence intervals vary by the age of the child and the specific assessment. SE_Ms for each level of the Screens, skill area scores, and supplemental assessments appear in the appendices within the table appropriate for each child’s age.

### Table 8-2. Internal Consistency of the Self-help and Social-Emotional Scales (Two-Year-Old Child and Higher), and the Readiness for Reading Scale (Five-Year-Old Child/Kindergarten)

<table>
<thead>
<tr>
<th>Basic Assessments</th>
<th>SELF-HELP SCALES</th>
<th>SOCIAL-EMOTIONAL SCALES</th>
<th>READINESS FOR READING SCALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guttman Lambda Coefficient</td>
<td>.88</td>
<td>.92</td>
<td>.89</td>
</tr>
<tr>
<td>Standard Error of Measurement</td>
<td>1.42</td>
<td>2.69</td>
<td>1.30</td>
</tr>
</tbody>
</table>
CHAPTER 9. Validity of the BRIGANCE® Screens

Critical Concepts in Screening and Test Construction: Validity

- **Content validity** answers the question, Do BRIGANCE® Screens items sample readiness skills considered to be important by developmental researchers and educators?

- **Construct validity** answers the question, To what extent do BRIGANCE Screens assessments measure unique dimensions of developmental and readiness skills?

- **Concurrent validity** answers the question, Do the BRIGANCE Screens relate (correlate) with more comprehensive measures (such as tests of intelligence, achievement, language, etc.)? Similarly, concurrent validity tells by correlations with other tests whether, for example, the BRIGANCE Screens language items correlate with other language tests or whether their academic items correlate with tests of achievement.

- **Predictive validity** answers the question, Does the test given at the beginning of the school year predict future performance, such as at the end of the school year?

- **Discriminant validity** answers the question, Are there unique patterns of performance on the BRIGANCE Screens that characterize children with developmental strengths or weaknesses?

- **Criterion-related validity** answers the question, How well do the BRIGANCE Screens identify children with and without developmental problems?

Content Validity

**DEVELOPMENT AND ORIGINAL STANDARDIZATION OF THE BRIGANCE SCREENS.** There is abundant support for the content validity of the BRIGANCE Screens and for the applicability of the Screens in educational settings. The construction of the BRIGANCE Screens was based on the authors’ extensive reading of developmental and readiness literature and on collaboration with numerous other educators who assisted in item selection. Other researchers compared various screening-instrument features such as ease and length of administration and whether item content reflects curricular content. For example, Helfeldt (1984) concluded that “the BRIGANCE Screens is a well-organized criterion-referenced test designed to assist in the early identification of individuals who need further testing.” Brennan (1985) commented that the BRIGANCE Screens has a high degree of “content validity due to its similarity to other well-established measures and has widespread use by professionals in the field of screening” (p. 4). Schearer (1986) interviewed 263 teachers, testing coordinators, and curriculum supervisors in New York City Public School districts, all of whom
Predictive Validity

Over the last two decades, the BRIGANCE® Screens have enjoyed tremendous positive professional scrutiny from researchers studying their validity and factor structure. Such studies have included invaluable longitudinal predictive validity research showing that a problematic BRIGANCE Screen at Kindergarten predicted difficulties with academic achievement in late elementary school (McClure & Benson, 1984; Wenner, 1988; Ellwein, 1991; McCarthy, 1994; Mantzicopoulos, 1999b).

Summary of Validity Research

- The BRIGANCE® Screens have substantial content validity. Items selected by a pool of educators and psychologists were drawn from research and other measures. The Screens also demonstrate desirable age-related trends: younger children average lower scores than do older children administered the same subtests.

- Construct validity is substantial. The BRIGANCE Screens contain a clear factor/skill area structure that accounts for the majority of variance within scores and with minimal overlap among factors.

- The BRIGANCE Screens have excellent concurrent validity. Skill areas, assessments, and factors are highly correlated with diagnostic measures of development, including language, preacademics/academics, intelligence, motor, social-emotional, and adaptive behavior.

- The BRIGANCE Screens have historically had substantial predictive validity and thus identified the majority of children who have school difficulty five months to six years later. Because the Screens are quite recent, new predictive validity studies need to be conducted. Nevertheless, since there is much overlap between the original and new versions of the Screens, we anticipate a high degree of predictive validity.

- The BRIGANCE Screens have a high degree of discriminant validity. Skill areas and factors correlate highly with developmental status and there are significant differences in the performance of children with typical versus atypical development, gifted versus non-gifted development, and between those with and without risk factors.
Comparing the BRIGANCE® Screens with Other Screening Tests

Figure E-1 illustrates a comparison of approximately fifty different screening tests. Tests were categorized and/or rated on (1) their content; (2) the measurement method (e.g., group, individual elicitation, parent report, observation, or combination); (3) the quality of standardization, reliability, and validity, and by their accuracy in terms of sensitivity and specificity; (4) the types of scores produced; (5) the types of personnel qualified to administer the measures; and (6) miscellaneous features, training time, administration time including scoring, and the age range of the test. Highlights from the table are presented below:

Better Measures

Of the seven best measures, three rely exclusively on parental report (the Child Development Inventories, Parents' Evaluations of Developmental Status, and the Ages and Stages Questionnaire) and are best suited for intake, mass mailouts, or screening in medical settings where children are less likely to cooperate with direct testing.

The Battelle Developmental Inventory Screening Test–II includes a heavy sampling of motor, self-help, and socialization items and a very limited sample of academic skills (i.e., there are no items tapping reading or reading readiness). It is lengthy, challenging to administer, and requires extensive training (Glascoe and Byrne 1993). For these reasons, it may be best suited for early intervention specialists who may benefit from its close ties with the larger Battelle Developmental Inventory, a diagnostic measure of development with utility in constructing IEPs.

The FirstSTEP is an instrument that, although it appears highly accurate, has not been subjected to research other than that conducted by the author. Further, its validation excluded children with known disabilities (which may result in an inflated estimate of typical performance) and then compared the standardization sample to an excessively large number of children with disabilities (30%), a situation that is known to spuriously inflate sensitivity and specificity. Further, it is not apparent that at-risk students (e.g., those who receive standard scores below 85 on measures of ability or achievement) were included in FirstSTEP’s validation and standardization. Finally, the FirstSTEP included a preponderance of highly educated families in its standardization sample.

The DIAL–III, although quite appealing to children, has been criticized for its limited sensitivity in detecting children with learning disabilities (57%) which is likely due to its relative absence of academic tasks (Jacob, Snider, and Wilson 1988). The recent revision did not thoroughly view its sensitivity and specificity—the most critical information determining the quality of a screening test. Although the DIAL–III may be accurate, this remains to be proven.

The Early Screening Inventory and the Preschool Screening System have desirable levels of sensitivity and specificity but poor standardization since both have only regional norms. This makes the national use of these measures questionable.
Accordingly, the BRIGANCE® Screens stand out as the logical choice for educational settings. Their use of criterion-referenced items, selected by professional educators together with norm-referenced scoring for comparing children to their peers, make the BRIGANCE Screens highly attractive to teachers and evaluation personnel. The link between the BRIGANCE Screens and the BRIGANCE Early Childhood/Head Start Developmental Inventories or the BRIGANCE Diagnostic Inventory of Early Development II is also a desirable feature for use in program planning. Above all, the BRIGANCE Screens have a high degree of accuracy and correctly classify 75% of at-risk children and 82% of typically developing children.

Less-than-Ideal Measures

Of the many poorly constructed measures, three have tremendous popularity and deserve discussion here. The Denver-II, although it has a high degree of sensitivity (if questionable and abnormal scores are collapsed), has extremely poor specificity (44%). For every six administrations, four children will fail; and of the four, only one will be found on diagnostic testing to have true difficulties (Glascoe et al. 1992). This makes the Denver-II expensive to administer since it will lead to a substantial number of referrals for unnecessary psychoeducational and language evaluations. The alternative method for scoring the Denver-II (collapsing questionable and normal scores) results in another problematic solution: sensitivity of only 44% and specificity of 85%. These less than parsimonious results may be due to the Denver-II's explicit exclusion of items measuring academic skills and lack of concurrent validation.

Despite its excellent conceptualization, standardization, sampling, stratification, and reliability analyses, the validity of the Early Screening Profile (ESP) is not supported, even by the authors’ own research. This may be due to the use of items discarded during the development of other tests written by the authors and items which generally fail to sample realistic behaviors. As a consequence, the ESP identifies fewer than 68% of children with difficulties and achieves this level of accuracy only for the (fairly lengthy) test as a whole. The subtests, particularly language, detect fewer than 40% of children with language problems.

Another popular but highly questionable measure is the Gesell School Readiness Test. Normed exclusively in New Haven, Connecticut, on upper socioeconomic-status White children, the test has limited test-retest agreement (73%) and poor inter-rater reliability (.71). This appears to be due to ambiguities in directions and scoring as well as to the improbability of acquiring the extensive training (seventy to seventy-five hours) needed to administer and interpret the test correctly. In a summary of Gesell research presented by Gredler (1992), between 53% and 62% of children who failed the Gesell were found to be quite successful in kindergarten. The Kaufman revision of the Gesell improved matters substantially and in a small study of 76 children, had excellent specificity and sensitivity. However, predictive specificity continued to be limited since 33% of children who performed poorly on the Gesell performed adequately on achievement measures by the end of first grade. Gredler summarized his review by noting “the percentage of error in labeling children as immature who in fact succeed is too large to continue use of the test in making placement decisions” (1992, p. 72).
<table>
<thead>
<tr>
<th>Domains Evaluated</th>
<th>Method of Measurement</th>
<th>Test Standardization</th>
<th>Type of Score</th>
<th>Examiner Qualifications</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross Motor Skills</td>
<td>Direct Observation</td>
<td>Standardization/Sampling</td>
<td>Age Equivalent</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Fine Motor Skills</td>
<td>Observation of Child</td>
<td>Reliability</td>
<td>None</td>
<td>Specialist</td>
<td></td>
</tr>
<tr>
<td>Social Skills</td>
<td>Group Administration Mass Matrif</td>
<td>Invalidity</td>
<td>None</td>
<td>Language Current/Nonsexist</td>
<td></td>
</tr>
<tr>
<td>Self-Help Skills</td>
<td>Direct Observation</td>
<td>Validity</td>
<td>None</td>
<td>Language CURRENT/Nonsexist</td>
<td></td>
</tr>
<tr>
<td>Cognitive Skills</td>
<td>Observation of Child</td>
<td>Sensitivity</td>
<td>None</td>
<td>Language CURRENT/Nonsexist</td>
<td></td>
</tr>
<tr>
<td>Phonological Skills</td>
<td>Direct Observation</td>
<td>Specificity</td>
<td>None</td>
<td>Language CURRENT/Nonsexist</td>
<td></td>
</tr>
<tr>
<td>Academic-Preacademic Language</td>
<td>Direct Observation</td>
<td>Standardization/Sampling</td>
<td>Age Equivalent</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Expressive Language</td>
<td>Observation of Child</td>
<td>Reliability</td>
<td>None</td>
<td>Specialist</td>
<td></td>
</tr>
<tr>
<td>Receptive Language</td>
<td>Group Administration Mass Matrif</td>
<td>Invalidity</td>
<td>None</td>
<td>Language CURRENT/Nonsexist</td>
<td></td>
</tr>
<tr>
<td>Articulation Skills</td>
<td>Direct Observation</td>
<td>Validity</td>
<td>None</td>
<td>Language CURRENT/Nonsexist</td>
<td></td>
</tr>
<tr>
<td>Behavior/Self Control</td>
<td>Observation of Child</td>
<td>Sensitivity</td>
<td>None</td>
<td>Language CURRENT/Nonsexist</td>
<td></td>
</tr>
<tr>
<td>Group Administration Mass Matrif</td>
<td>Direct Observation</td>
<td>Standardization/Sampling</td>
<td>Age Equivalent</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Direct Observation</td>
<td>Reliability</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation of Child</td>
<td>Validity</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History/Interview</td>
<td>Sensitivity</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardization/Sampling</td>
<td>Specificity</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner Qualifications</td>
<td>Language CURRENT/Nonsexist</td>
<td>None</td>
<td>None</td>
<td>Language CURRENT/Nonsexist</td>
<td></td>
</tr>
</tbody>
</table>

More Recommended

- BRIGANCE® Screens
- Parent’s Evaluations of Developmental Status (PEDS)
- Ages and Stages
- Battelle Inventory-Screen (BDIST-II)
- Bayley Infant Neurodevelopmental Screener (BINS)
- FirstSTEP
- DVL-III
- SCREEN
- Early Screening Inventory
- Preschool Screening System
- Developmental Profile II
- Dallas Preschool Screening Test
- Bosan-Inelligence Test
- Diaberno
- Denver II
- Denver Developmental Screening Test (Revised)
- Early Screening Profile (ESP)
- Denver Prescreening Developmental Questionnaire
- Gesell Preschool & School Readiness Tests
- Lexington Developmental Scales (Short Form)
- Cooperative Preschool Inventory
- Comprehension Identification Process
- Early Detection Inventory
- The ABC Inventory
- The Quick Test
- ANSIER
- Rile Preschool Screening Test
- Birth to Three

Less Recommended

Adapted from “Pediatrics,” Glascoe, F.P., et al., 1990, 86: 547-554 and the newsletter of the National Association of School Psychologists

Compilled by Frances Page Glascoc, Ph.D., Child Development Center Department of Pediatrics, Vanderbilt University

LEGEND: X = Present or Yes  E = Excellent or 80%  G = Good or 70% – 79%  F = Fair or 60% – 69%  P = Poor or < 60%  -- = Absent