



Executive summary: GDO Study Reliability and Validity Evidence

Version: June 19, 2012

Author: Gesell Institute of Child Development

Abstract: This document presents, in executive summary format, reliability and validity evidence collected in the GDO Study to support the continued use of the Gesell Developmental Observation-Revised and the new Gesell Early Screener with children age 3-6.

Keywords: Child development, developmental assessment, GDO Study, GDO-R, GES, technical data, early childhood assessment, PreK assessment, Gesell

## OVERVIEW OF THE GESELL DEVELOPMENTAL OBSERVATION-REVISED

The Gesell Developmental Observation-Revised (GDO-R) is a standardized, performance-based, criterion-referenced instrument that measures a child’s behavior in four domains through direct observation, and by surveying parents and teachers. The GDO-R provides a Developmental Age and Performance Level Ratings in five strands.

The GDO-R is designed to monitor development of children ages 2 ½ to 9 over years and grades. Examiner training is required. The performance-based results may be used to analyze a child’s developmental and learning profile, to strategize for curriculum planning, and to report progress to parents. The GDO-R also indicates whether or not a child may need further evaluation in specific domains of development (Cognitive, Language, Motor, or Social-Emotional). Refer to Table 1 for the domains measured by the five strands.

Table 1  
*Domains Measured by GDO-R Strands*

Domain/ Strand	Strand A	Strand B	Strand C	Strand D	Strand E
	Developmental	Letters/ Numbers	Language/ Comprehension	Visual/Spatial Discrimination	Social/ Emotional/ Adaptive
Cognitive	✓	✓	✓	✓	
Language	✓	✓	✓	✓	✓
Motor	✓	✓		✓	✓
Social/ Emotional/ Adaptive	✓		✓		✓

GDO-R tasks are divided into five strands (sets of related tasks). Scores from each strand result in the following Performance Level Ratings:

- **Age Appropriate** (solid or qualified expectation responses for all or most of the tasks in the strand)
- **Emerging** (solid or qualified expectation responses for most or only some tasks in the strand)
- **Concern** (atypical responses for most tasks in the strand).

## PURPOSE OF THE STUDY

The 2007 Gesell Developmental Observation (GDO) was the focus of a nationwide study of 3-6 year old children from 2008-2010 (GDO Study). The first purpose of the GDO Study was to provide updated technical data and validity evidence for 17 of the 19 tasks on the ©2007 GDO to support its continued use as a comprehensive developmental assessment system. The second purpose was to comply with current state and federal standards which mandate that schools use child assessments which meet the psychometric standards set forth by the American Psychological Association. The third purpose of the GDO Study was to publish a complete revision of the instrument, including a new examiner's manual with corresponding forms and surveys.

Data was collected from children (child developmental and academic assessment data-GDO), teachers (teacher survey of child behavior, social interactions, preferences and adaptive skills in the classroom-TQ), and parents (parent survey of family characteristics, experiences, child behavior, and adaptive skills at home-PQ). In addition, data was collected on child pilot items to assess growth patterns of numeracy, literacy, and social-emotional interactions. Examiner demographics (sex, ethnicity, level of education and GDO experience) and school population demographics (child ethnicity, sex, age and eligibility for the federal Free/Reduced Lunch program) were also collected to better understand the wider sample from which the GDO assessments originated.

## GDO STUDY PREPARATION

Prior to collecting data, the research team conducted a series of investigations to provide content validity for the instrument. Content-related validity is evidenced by uniformity between strand and task content and the developmental milestones widely accepted to precede instructional content in each area. To ensure such correspondence for the revised GDO, Gesell Institute conducted a comprehensive review of current child development theory, and met with education experts to determine common educational goals and the knowledge and skills emphasized in today's early childhood curricula.

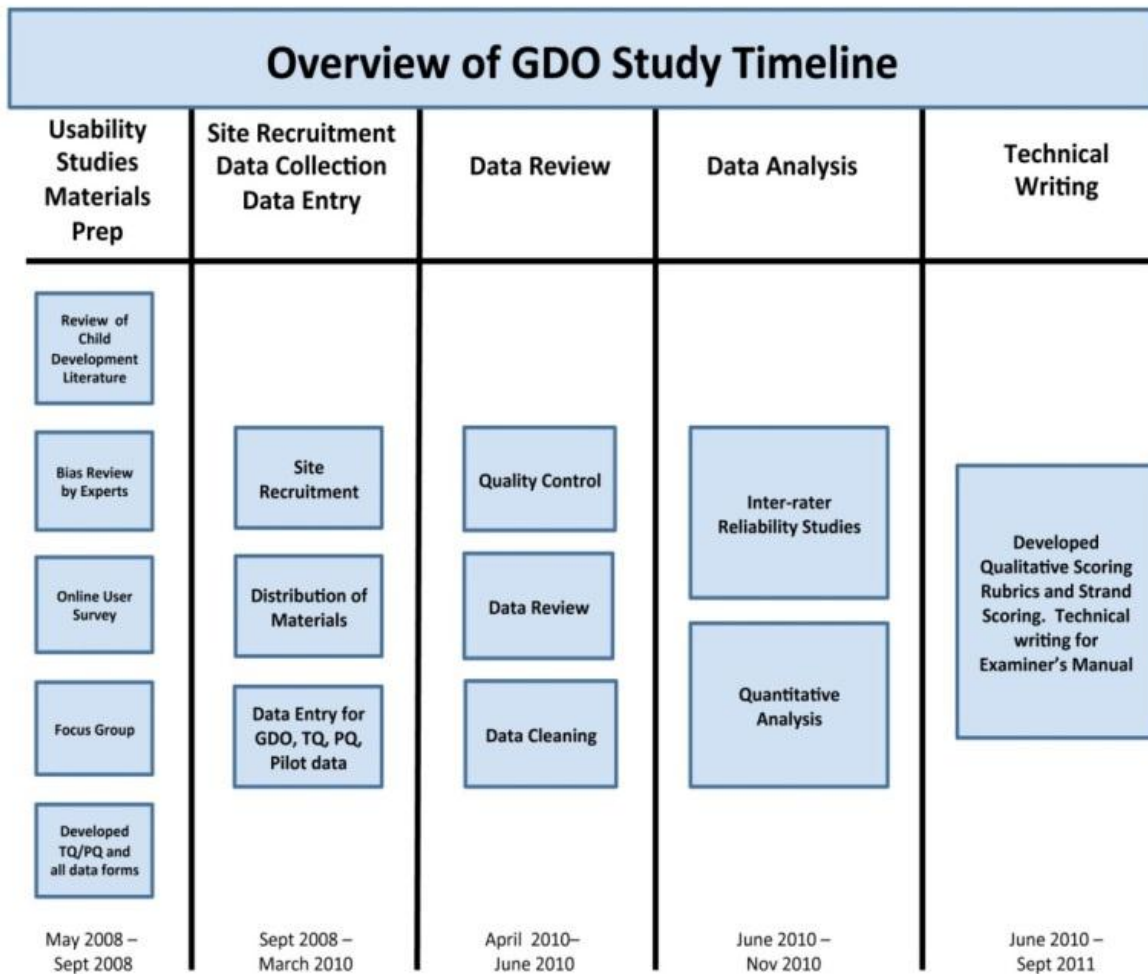
The GDO was reviewed by a group of five experts from the fields of early childhood education, special education, physical movement, and test development. Each reviewer was individually asked to respond to a set of specific questions, in order to identify any biases inherent in the content or methodology of the GDO. It was recommended that the revised GDO include a measure of emotional regulation and social behavior by surveying the child's teacher and/or parent, and to improve consistency in administration across tasks. A Teacher Questionnaire (TQ) and Parent/Guardian Questionnaire (PQ) were developed for the GDO Study, as well as a revised Examiner's Script.

An on-line user survey and focus group provided additional information regarding overall assessment effectiveness (addressing such topics as the criterion for assigning developmental age, and appropriateness for each task by age). These validation efforts resulted in a streamlined assessment that reflects the needs of classroom teachers, children, and parents. Other modifications resulting from the GDO Study included objective scoring rubrics, new strand scoring, Performance Level Ratings, and improved item language.

Data was collected on each item of the designated tasks and on the strand level to measure the psychometric properties of the instrument. Analyses included descriptive statistics, correlations, and inter-rater reliability findings for Developmental Age. The comprehensive Technical Report contains quantitative and qualitative evidence for reliability and validity.

The GDO Study consisted of multiple phases of activities. Participating schools were recruited from a sample of current GDO users across the U.S. Schools joined the study at various points in time, and collected data at varying rates. Refer to Figure 1 below.

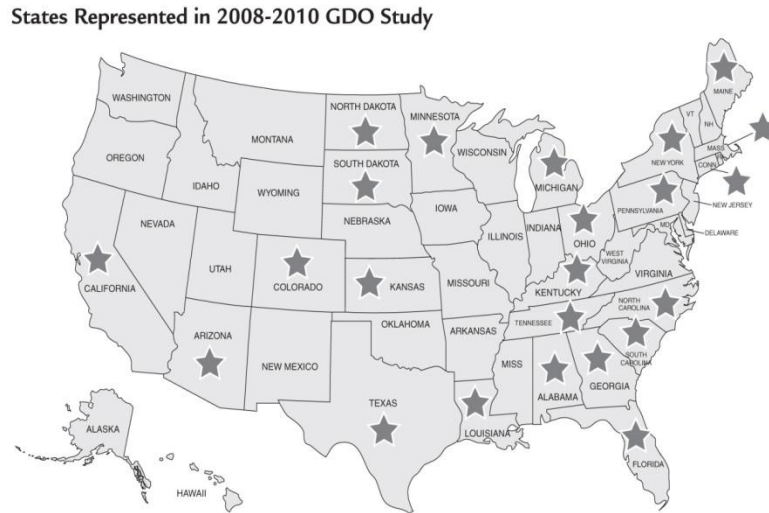
Figure 1



## GDO SAMPLE DEMOGRAPHICS

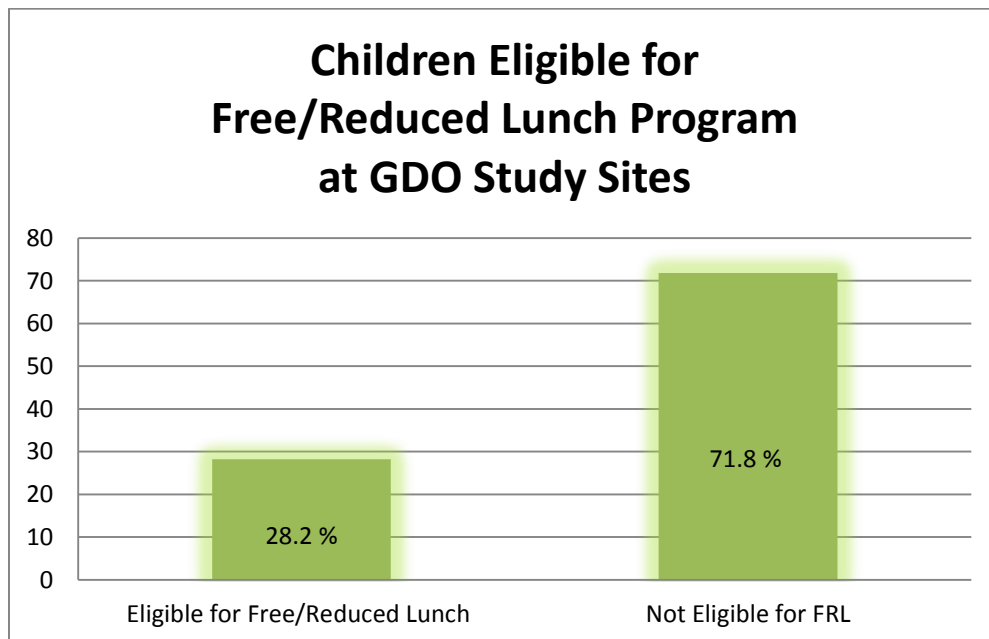
The GDO Study collected data from a nationwide, culturally diverse sample of 1,287 children age 3-6 located in 53 sites spanning 23 states. Refer to Figure 2 below.

Figure 2



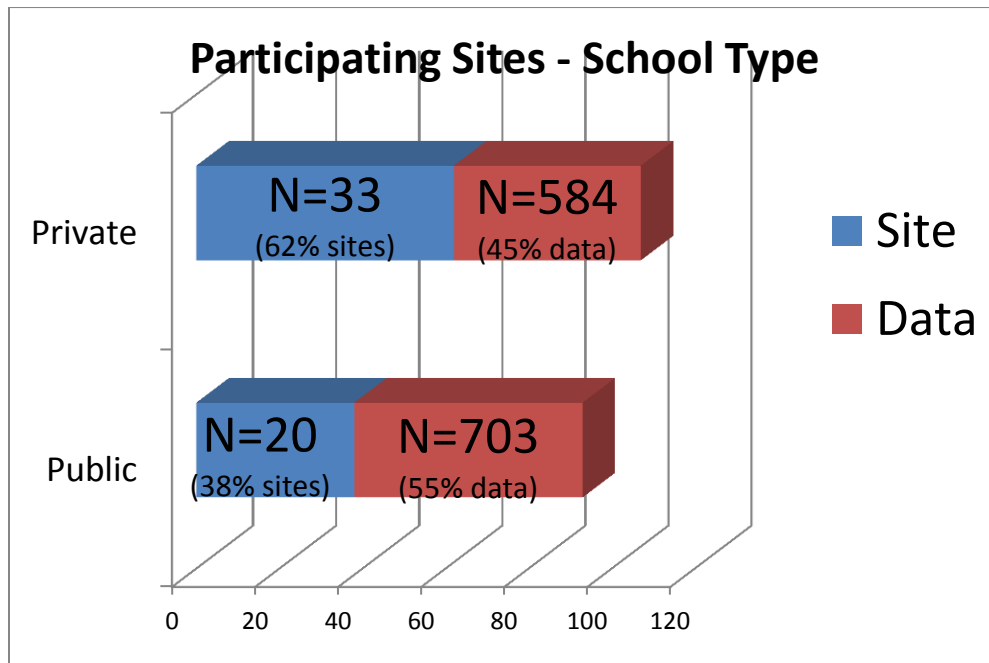
Each site was surveyed for the number of children between the ages of 3-6 who were eligible for the federal Free/Reduced Lunch (FRL) program in their school. The sample of children eligible for FRL at each school site is a representative measure for the children in the GDO Study sample with similar eligibility. Refer to Chart 1 below.

Chart 1



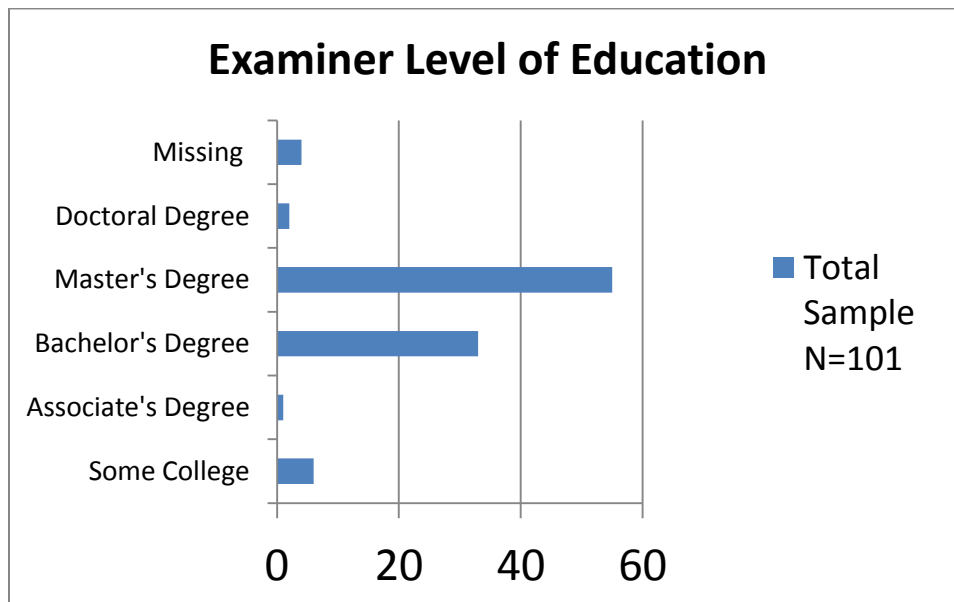
A greater number of private schools participated in the GDO Study than did public schools. However, a greater number of child GDO assessment data was submitted from public school sites than from private school sites. Refer to Chart 2 below.

Chart 2



The group of 101 GDO volunteer examiners had a mean of 7 years GDO administration experience. All examiners received GDO training within the past 5 years, and also received current standardized materials and individual consultation from research staff at Gesell Institute. Eighty-eight percent of the GDO examiners hold a Bachelor's, Master's, or Doctoral degree. Refer to Chart 3 below.

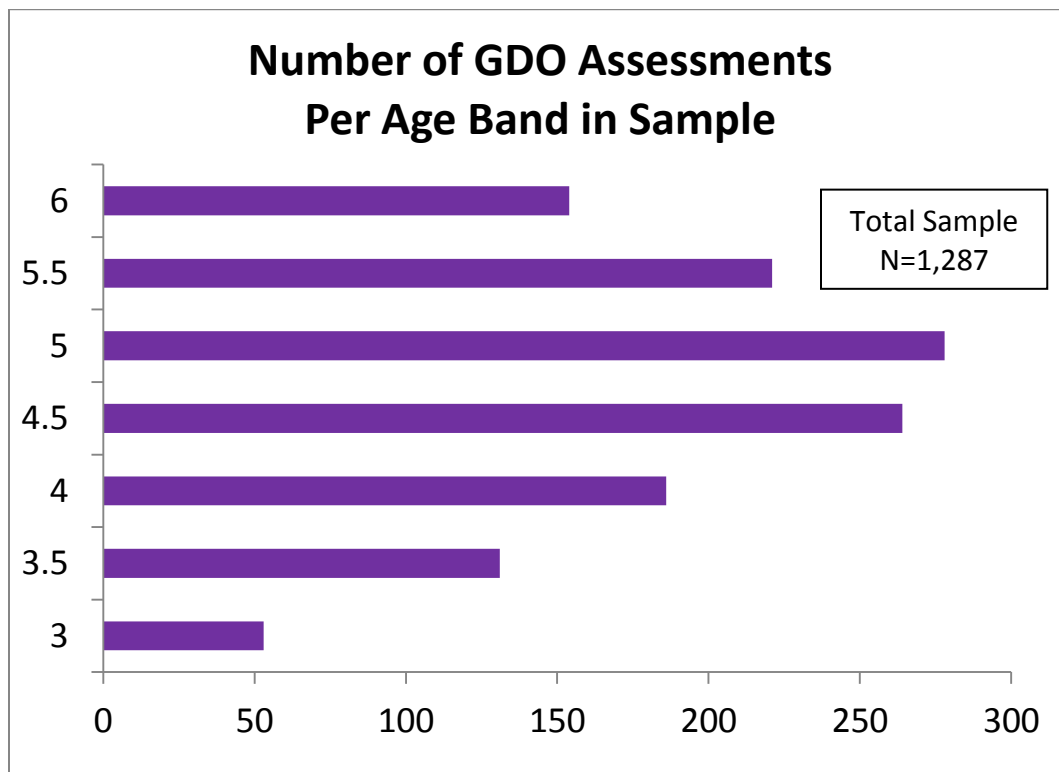
Chart 3



A target of 200 cases per age band was set prior to commencing the GDO Study. The target sample was equally distributed in 6 month age bands from 2 years 4 months to 6 years 3 months. Thus, each age band included child assessment data from children who ranged from 3 months younger than to three months older than the age band. The 2 ½ age band was eliminated due to insufficient data. The number of cases in the 3, 3 ½ and 6 year age bands was low (53, 131 and 154 respectively), but all other age bands met or exceeded the goal.

Children between 7 and 9 years were not included in the study because this assessment is not used primarily with this age group. However, new information is provided in the Examiner’s Manual for the 2 ½ age band and for older children (6 year 4 months to 9 years old) based on the knowledge of a group of expert examiners. Refer to Chart 4 below.

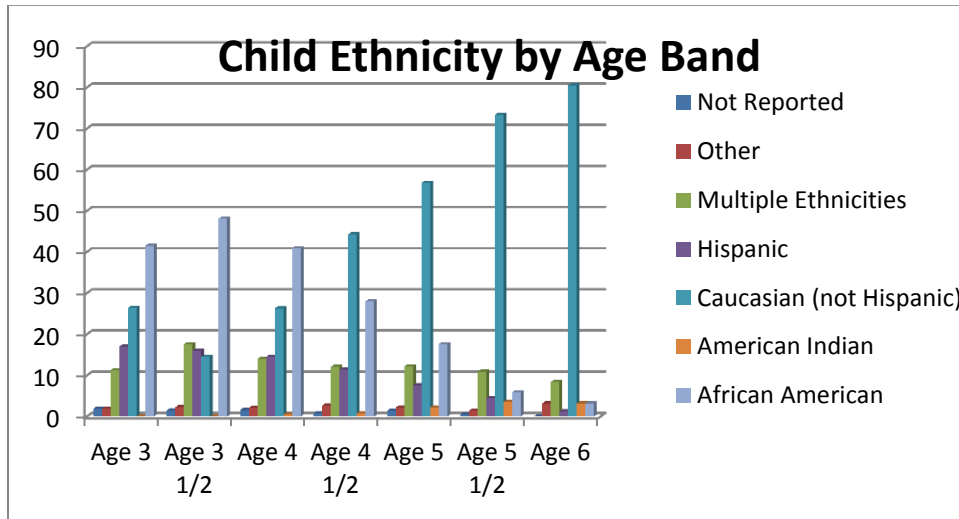
Chart 4



Age Band	3.0	3.5	4.0	4.5	5.0	5.5	6.0
Number of Students	53	131	186	264	278	221	154

The following chart describes child ethnicity by age band, and indicates that more Caucasian children were in the 4 ½ - 6 year bands and more African American children were in the 3-4 year bands. Refer to Chart 5 below.

Chart 5



Age Band	3.0	3.5	4.0	4.5	5.0	5.5	6.0
Ethnicity (%)							
African-American	41.5	48.1	40.9	28.0	17.6	5.9	3.2
American Indian	0.0	0.0	0.5	0.8	2.2	3.6	3.2
Caucasian not Hispanic	26.4	14.5	26.3	44.3	56.8	73.3	80.5
Hispanic	17.0	16.0	14.5	11.4	7.6	4.5	1.3
Multiple Ethnicities	11.3	17.6	14.0	12.1	12.2	10.9	8.4
Other	1.9	2.3	2.1	2.7	2.2	1.4	3.2
Not Reported	1.9	1.5	1.6	0.8	1.4	.5	0.0

## RELIABILITY AND VALIDITY EVIDENCE

Analyses of the technical data collected in the 2008-2010 GDO Study support the developmental age expectations for the tasks originally published by Dr. Arnold Gesell over the last six decades. Data originally published by Gesell utilized a modest 51% “benchmark” for identifying developmental competence on a task. However, today, the GDO-R employs a more contemporary framework for performance on each task which utilizes three Performance Level Expectations, rather than two. These are Solid (over 70% of all children performed task), Qualified (50-69% of children performed task) and Not Yet Expected (less than 50% performed task). Moreover, scoring rubrics for a child’s performance on each strand of tasks further extends the utility of the instrument by ultimately defining three Performance Level Ratings: Age Appropriate, Emerging and Concern. This strand scoring is a valuable scoring feature of the revised GDO and was derived from multiple sources: GDO Study data, the judgment of expert examiners, and established theories of child development.



The GDO Study measured the extent to which highly trained examiners could reliably assign a Developmental Age score to samples of children's Copy Forms and Incomplete Man work. Expert trainers (Gesell Institute National Lecture Staff) achieved very high inter-rater reliability correlations (.91-.93), and thereby re-established the reliability of the qualitative training provided for the instrument.

## LIMITATIONS

Several limitations exist with the present study. The first is that the two youngest age bands (3 and 3 ½) and the oldest age band (6) contained fewer child GDO assessments than were targeted. The second is that the distribution of child ethnicity across the total sample more closely approximated the U.S. Census than did the distribution within each discrete age band. This is likely the result of a third limitation, having sampled from schools that were defined as trained GDO users (i.e., sample of convenience) as opposed to sampling schools that were defined according to a certain demographic profile. Lastly, GDO examiners did not receive rubrics to score certain qualitative items such as Paper Position, Head Shift, Body Posture, Non-dominant Hand Posture, or Eye Movement. Thus, data for these items cannot be clearly interpreted.

## IMPORTANT IMPROVEMENTS

The GDO-R provides several important new features that enhance the usability and reliability of the instrument as a whole over the 2007 GDO. These are strand scoring for five strands which results in three Performance Levels, electronically facilitated scoring, and the first comprehensive Technical Report. The GDO Study also resulted in the development of the Gesell Early Screener, a complementary short screening instrument which requires no training, flags children in need of further evaluation in any of four domains, and utilizes the TQ and PQ measures as well.

## CONCLUSIONS

The GDO Study contributes a comprehensive body of child development data to the educational field at large. This data authenticates the central role child development plays in understanding how children learn and in creating appropriate early learning experiences for children 2 ½-9. The data demonstrates the stability of growth through ages and stages of development, and confirms Dr. Gesell's original findings. The data also supports current brain research which indicates that pathways for new information are neurologically connected to and built upon previously established synapses. Inter-rater reliability evidence for determining Developmental Age endorses the renewed use of GDO-R, and the developmental characteristics for Copy Forms and Incomplete Man tasks in particular.

The results of the GDO Study can be effectively generalized to a wider population of young children because the study sample utilized experienced examiners trained on standardized GDO administration, operationalized discrete age bands for comparison of data, and included children from socioeconomically, demographically, and culturally diverse backgrounds.